New 3D measurements of forest structure

Andrew Philip Burt

Department of Geography

University College London

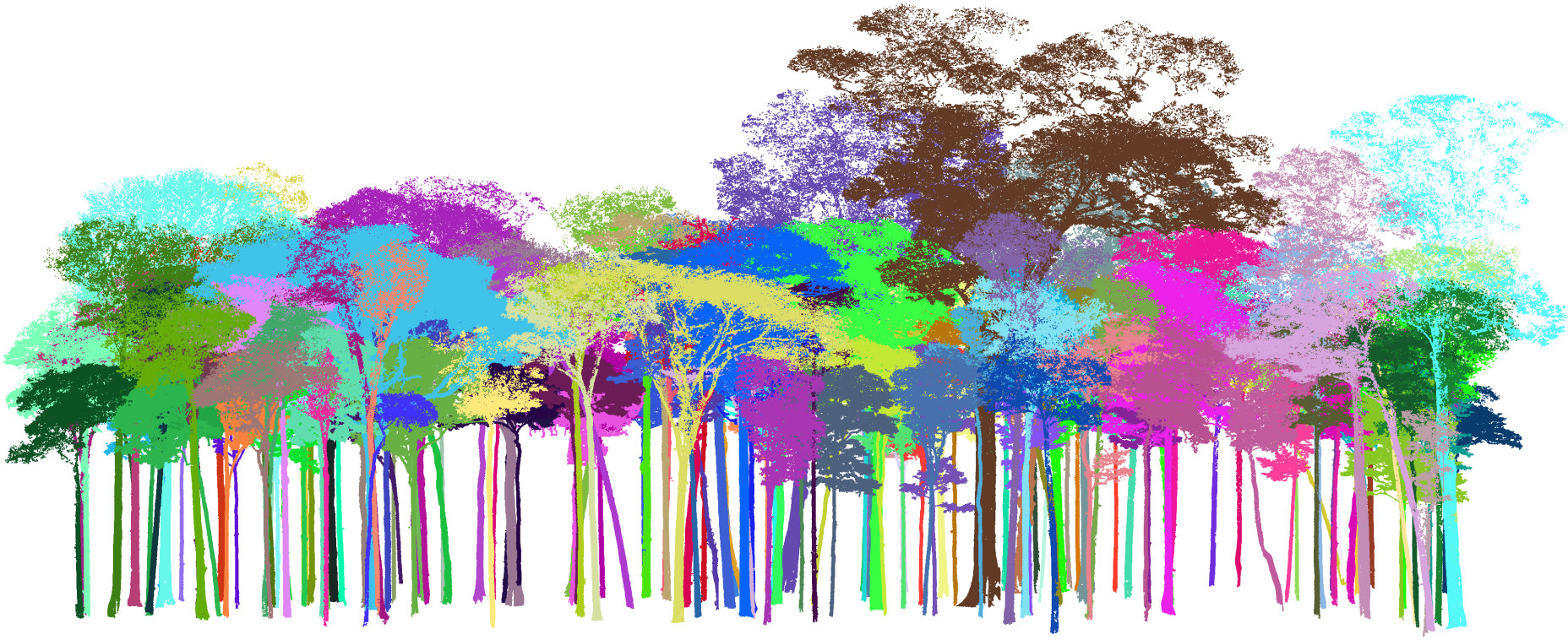
July 2017

Submitted for the degree of Doctor of Philosophy

Supervised by Dr. Mathias Disney

I, Andrew Burt, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

.....



Individual tree-level point clouds derived from terrestrial laser scanning data acquired in the tropical forests of Caixuanã National Forest, Brazil.

Abstract

Regular assessment of the state and change of the world's forests is essential because of the range of climate and ecosystem services they provide. Earth observation efforts responsible for monitoring the above-ground biomass (AGB) of the world's forests, a frequently used proxy for forest state, rely on calibration from a network of field plots. In lieu of impracticable direct measurement of AGB at these plots, estimates are obtained through empirical allometric correlations that relate simple measurements of tree structure, to mass.

Throughout tropical forest, few pan-tropical allometric models exist because of the limited availability of calibration data. In this thesis, the best available pan-tropical allometric dataset is used to demonstrate that the statistical method employed in the construction of all widely used models, log-transformed ordinary least squares linear regression, is inappropriate.

Alternatively, a new non-linear model is proposed where uncertainties are derived from non-parametric methods. These uncertainties are shown to exceed 75 % and 25 % of AGB at the tree- and plot-scale respectively. This leads to the conclusion that AGB estimates of large swathes of tropical forest are statistically indistinguishable from one another when inferred through pan-tropical allometry.

These results highlight the need to introduce alternative methods. The primary objective of this thesis is to demonstrate that new 3D measurements of forest structure from terrestrial laser scanning (TLS) can provide accurate non-destructive estimates of AGB. Rich point clouds have been acquired across multiple forest types. Novel algorithms are developed and applied here to retrieve tree-scale AGB via volume estimation on a wide-scale.

It is shown these new methods can estimate tree- and plot-scale AGB to within 23.3 % and 7.9 % of the respective direct measurement. It is shown these TLS-derived estimates of AGB, are, on average, 17.9 % larger than their allometric-derived counterparts. Allometric methods may thus significantly underestimate tropical forest carbon stocks.

Contents

Acknowledgements

I would like to express my gratitude to my supervisor, Mathias Disney, for giving me the opportunity to undertake this PhD. Thank you for your guidance and encouragement, and for imparting your knowledge and enthusiasm of the natural world to me.

I am indebted to a number of people who made this research both possible and enjoyable. In particular, I wish to thank my secondary supervisors, Philip Lewis and Simon Lewis; as well as Kim Calders, Jose Gómez-Dans, Pasi Raumonen, John Armston, Aida Cuni Sanchez and the members of both TLSIIG and TLSRCN. I acknowledge Forest Research and Eric Casella for the Alice Holt Forest destructive harvest campaign, and the resulting data.

I would like to thank my mother, Phyllis, for her love and encouragement. Finally, I would like to thank my father, Geoff, for his inspiration and endless support.

This research was funded by the Natural Environment Research Council through grant NE/J016926/1.

1 Introduction

1.1 The global role of forests

Forests play a key role in sustaining life on earth. Alongside ocean phytoplankton, the world's forests are jointly responsible for primary production, and its by-product, the release of oxygen into the atmosphere (Field et al. 1998). Both regionally and globally, climate patterns are regulated by the world's forests (Bonan 2008; Jasechko et al. 2013), and they additionally balance regional ecosystems through their modulation of water flow (Bruijnzeel 2004).

The socio-economic benefits of the world's forests are considerable, with somewhere in the region of 1.6 billion livelihoods directly dependant on forest resource harvesting (The World Bank 2004). Formally, the forestry sector contributes 0.9 % to global GDP and employs 50 million people worldwide (FAO 2016). In more commodity-linked, forest-rich markets, such as Ghana, this contribution by the forestry sector to total GDP is as large as 3.5 % (Lebedys et al. 2014). It has been estimated that the total value of global forest ecosystem services is $\$16.2\,\text{trillion}\,\text{yr}^{-1}$ (Costanza et al. 2014).

The world's forests also provide habitat for approximately 80 % of terrestrial biodiversity (FAO 2012). Harboured in the world's forests are a number of endangered animals and plants, many of which are either keystone species, or provide many wider ecosystem benefits ranging from pollination through to medicine (Richards 1952; Bawa 1990).

Forests also play a key role in the global carbon cycle. An estimated sink of $1.1\,\text{Pg}\,\text{C}\,\text{yr}^{-1}$ from the atmosphere to forested land cover dominates terrestrial carbon sequestration (Pan et al. 2011). This sink helps to constrain atmospheric carbon uptake from anthropogenic influences to $4.0\,\text{Pg}\,\text{C}\,\text{yr}^{-1}$ (Ciais et al. 2013). Achievable rates of afforestation and reforestation incentivised by intergovernmental programmes including REDD+ have the potential to increase this sink by $0.4\,\text{Pg}\,\text{C}\,\text{yr}^{-1}$ (Canadell et al. 2008), enabling forests to become more potent mitigants of climate change.

1.2 Monitoring the world's forests

Despite this broad basket of ecosystem services provided by the world's forests, they are endangered by, amongst other things, deforestation. In the early Holocene, and in particular, prior to industrialisation, the complexion of the world's forests would have been strikingly different to that of today. Across several centuries, anthropogenic deforestation has led to the clearing of roughly 3.0 billion ha of forest (Bryant et al. 1997), reducing global forested land cover down from around 60 %, to 30 % (FAO 2015).

Since the 1990s, when the systematic monitoring of the state and change of the world's forests was first introduced, net deforestation stood at around 7.3 million ha yr^{-1} (Williams 2003). Through an increasing global awareness of the importance of the world's forests, net deforestation reduced steadily through the 2000s to 4.0 million ha yr^{-1}, and down to a current rate of approximately 3.3 million ha yr^{-1} (FAO 2010; FAO 2015). That is, at the current rate of net deforestation, somewhere in the region of 0.08 % of the world's forests are lost every year. Coupling this current and historical loss with the aforementioned ecosystem services provided by the world's forests, the regular and accurate assessment of their state and change is essential.

Biomass, the mass of living and dead organic matter, particularly above-ground biomass (AGB), is frequently reported in the context of assessing forest state. This is principally due to the intertwined relationship of AGB with both primary production and carbon. During photosynthetic activity, a carbon sink from the atmosphere manifests as biomass accumulation. Inversely, beside respiration, the decay of dead biomass results in a carbon emission. This makes AGB a popular parameter for carbon accounting, as a relatively stable relationship exists, with the carbon content of AGB approximately 50 % (Brown 1997; Clark et al. 2001).

The current benchmark AGB map of the tropics is presented in figure 1.1. A map of tropical forest has been illustrated here because tropical forest is perhaps the most studied of the forest types. This is because of their hyper-biodiversity, with 96 % of all tree species (Fine et al. 2008), and somewhere over 50 % of all terrestrial biotic species, found in tropical forests (Pasquis et al. 2004). Tropical forests also account for in excess of 50 % of the world's forests by carbon mass, despite only accounting for about 33 % of forested land cover (IPCC 2000; Bonan 2008). Yet tropical forest is also some of

the most at risk of being deforested, chiefly because of their location throughout the developing world.

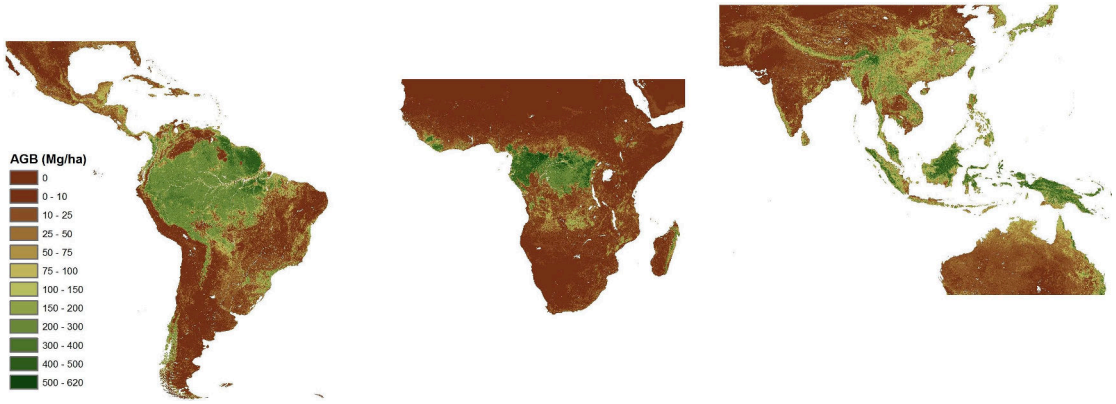


Figure 1.1: Pan-tropical AGB map (Source: Avitabile et al. 2016).

The ability to understand and quantify the global carbon cycle is reliant on such AGB products as that of figure 1.1 because of the contribution of forest ecosystems to terrestrial carbon sequestration. Similarly, any uncertainties associated with these AGB products will blur this ability (Houghton et al. 2009). Current 'best estimate' pan-tropical AGB maps carry significant uncertainty, with an expectation in the literature for them to routinely exceed 50 % of AGB (Saatchi et al. 2011). Furthermore, limited corroboration is discerned between independently derived maps based on the same data, with differences exceeding 200 $\mathrm{Mg\,ha^{-1}}$ across the full swath of the Amazon Basin (Mitchard et al. 2014). Most pan-tropical AGB maps, including that of figure 1.1, are derived from spacebourne earth observation, through the fusion of data from LiDAR, RADAR and multi-spectral optical instrumentation.

This uncertainty then, is introduced by two factors; firstly, the data analysis methods employed to extract the information content and account for the spatial and temporal limitations of the observations (Goetz et al. 2009). The other and frequently overlooked component is the calibration of the remotely sensed signal to a parameter that it is not directly measuring (Woodhouse et al. 2012). That is, these remote sensing methods are not measuring AGB, they are measuring the interaction of radiation with the earth's surface.

Routinely underpinning regional and wider-scale AGB estimates is a canopy height product (Lefsky et al. 2005). A correlation exists between canopy height and equivalent spatial-scale AGB, permitting AGB estimation through this canopy height product

(Baccini et al. 2012). This fundamental relationship, enabling large-scale estimation of AGB, is calibrated using a network of field plots, typically established at the hectare-scale (Malhi et al. 2002; Lewis et al. 2013), at which the parameters of tree height and AGB are characterised.

Although these plots are routinely referred to as 'ground truths', this would be a misnomer as the direct measurement of AGB is rarely made. Weighing, the direct measurement of AGB, through destructive harvest, of a representative tropical canopy tree, is a challenge. At the footprint size of spacebourne remote sensing instrumentation, across a series of plots, destructive harvest is impractical. Instead, AGB must be estimated through simple in-situ measurements of tree form (i.e., shape and size) from which AGB can be inferred through allometry.

1.3 Allometric estimation of tree- and plot-scale above-ground biomass

Allometry is the study of relationships between the form of an organism with respect to other biophysical parameters of the organism, including the proportionality of changes to such parameters in response to a relative change in form (Huxley et al. 1936). Here, woody plant life, such biophysical properties range from metabolic rate to above and below-ground biomass (Kittredge 1944).

The mechanics behind why tree form drives these relationships are vigorously contested. Classical theory starts from the premise that the primary determinant of tree form is the minimisation of the energy requirement for resource transportation through the body of the organism, here the rooting and branching system (West et al. 1997; Brown et al. 2004). Under such constraint, tree architecture, through a fractal-like framework, can be described by a power law, with a constant exponent of $\frac{3}{4}$, derived from the Navier-Stokes equations governing fluid flow (Landau et al. 1987). An increasing body of evidence rebuts this hypothesis, suggesting that environmental influences, in relatively undisturbed forest, such as light and nutrient competition take precedent in driving tree form (Kozlowski et al. 2004; Coomes et al. 2009; Coomes et al. 2011).

Few observations are required to identify that such relationships will broadly persist at the population-scale (i.e., the aforementioned correlation between tree height and AGB) but can be highly variable at the individual-scale. In order to capture this variability encountered in the natural world, allometric methods are nearly always derived

empirically, rather than from theory, often masking causation (Zianis et al. 2006).

Notwithstanding causation, these allometric relationships permit readily measurable parameters of tree form, such as bole diameter, D, tree height, H, and average wood density, ρ, to infer, other, harder to measure parameters, such as AGB, in the typical form:

$$AGB \propto (DH\rho)^{\beta} \tag{1.1}$$

Where β is some allometric exponent.

Across tropical forest, the most widely used allometric models (Brown 1997; Chave et al. 2005; Feldpausch et al. 2012; Chave et al. 2014) are constructed at the population-scale because of tree-scale variability (Gibbs et al. 2007). This implies that pan-tropical allometric models must represent somewhere in the region of 1.39 trillion trees (Crowther et al. 2015) transcending approximately 46,500 species (Slik et al. 2015).

The implication of uncertainties that arise from employing such models to infer tree- and plot-scale AGB are hard to overstate. Characterisation of the network of field plots responsible for calibration of all up-scaled AGB products, such as that shown across the tropical belt in figure 1.1, are wholly reliant on allometric modelling (Clark et al. 2012). So, it should be expected that uncertainty associated with allometric prediction of AGB will propagate directly and unabated into such products (i.e., the accuracy of regional-scale AGB estimates will not surpass the accuracy of plot-scale AGB estimates). Furthermore, implied through the UNFCCC designation of AGB as an Essential Climate Variable (GCOS 2010), subsequent products partially derived from AGB, such as global carbon stocks and fluxes, will, to some extent, be burdened by any uncertainty in the allometric estimates (Houghton 2005).

However, arising from the so-called 'fallacy of misplaced concreteness' in these field sites (Clark et al. 2012), uncertainty arising from the application of allometrics to characterise tree- and plot-scale AGB is routinely overlooked throughout the literature. The primary method for uncertainty quantification in pan-tropical allometric estimates of tree- and plot-scale AGB are those described by Chave et al. (2004), that when applied to the most widely used pan-tropical allometric models, estimate total uncertainty in the prediction of AGB to be approximately 55 % and between 5–10 % at the tree- and

1 ha plot-scale respectively.

1.4 Forest structure from terrestrial laser scanning

The primary objective of this thesis is to demonstrate that new methods derived from terrestrial laser scanning (TLS) are capable of providing non-destructive estimates of tree- and plot-scale AGB that surpass the accuracy of that permitted by the current methods of allometry.

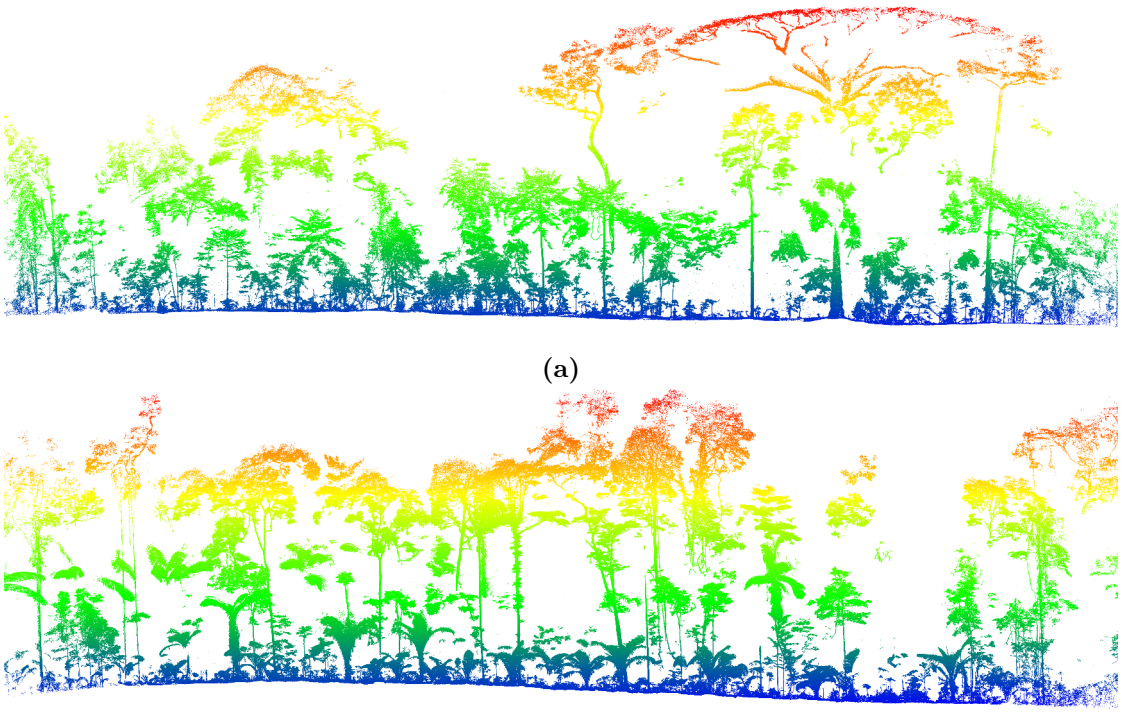
This thesis begins by presenting a novel experimental study critiquing the current methods in pan-tropical allometric modelling to determine the veracity of the current expectations of uncertainty in allometric-derived predictions of AGB. This study is presented from the outset because judging the efficacy of new TLS-derived methods will be dependant on understanding the biases and uncertainties that are present in the current 'best estimate' methods.

The base method of TLS, light detection and ranging (LiDAR), is the estimation of range between the instrument and some element in the forest scene through the time-of-flight of an emitted pulse of radiation. Acquiring multiple measurements, from many locations throughout the forest scene will produce a geometric sampling, that is realised through the construction of a plot-level point cloud.

In the work for this thesis, TLS has been deployed in the forests of Australia, Gabon, Brazil, the United Kingdom and French Guiana. Rich 3D point clouds, composed of more than 4 billion pts ha^{-1}, at 5 mm accuracy, have been captured. To illustrate these data, figure 1.2 presents an example of two slices from plot-level point clouds acquired in the tropical forests of Gabon and French Guiana.

Whilst the information content of these clouds is clearly abundant, access to it is non-trivial as the sole attribute of the data is the relative 3D arrangement of the points to one another. Combining the complexity of these data with the relatively recent emergence of sufficient quality commercially available instrumentation, only a small body of exploratory work has looked to retrieve forest structure from TLS data (Lovell et al. 2003; Watt et al. 2005; Maas et al. 2008; Côté et al. 2009; Calders et al. 2014).

More specifically, prior to commencing the work for this thesis, neither acquisition of TLS data in tropical forest, nor explicit quantification of TLS-derived tropical tree-scale AGB had ever been attempted. To retrieve tree-scale AGB from these data then, on a



(a)

(b)

Figure 1.2: Slices through plot-level point clouds obtained in the tropical forests of (a) Lopé National Park, Gabon and, (b) Nouragues Nature Reserve, French Guiana; both slices, coloured by height and with dimensions of 150 m x 2.5 m (LxW), contain approximately 100 million points.

wide-scale (here meaning tens of 1 ha forest plots), such that these new methods can be considered alternatives to allometrics, the following knowledge gaps must be bridged:

i) The automatic extraction of individual tree-level point clouds from plot-level point clouds.

ii) The automatic inference AGB from these individual tree-level point clouds.

iii) The quantification of uncertainty in these new TLS-derived estimates of AGB

Knowledge gap i) concerns the requirement to isolate tree-level point clouds from plot-level point clouds, such as those shown in figure 1.2, so as to permit access to the tree-specific information content of the TLS data. This thesis presents *treeseg*, an algorithm that has been developed here to achieve the near-automatic extraction of tree-level point clouds from plot-level point clouds. Examples of these tree-level point clouds are presented in figure 1.3, extracted from TLS data acquired from 0.25 ha *Eucalyptus* spp. open forest plot in Karawatha Forest Park, Australia.

From these tree-level point clouds, AGB can be inferred by fitting geometric prim-

itives to specific sections of the point cloud to reconstruct the surface topography of the tree. Methods such as the Raumonen et al. (2013) algorithm permit the construction of so-called 'quantitative structural models' (QSMs) that explicitly represent 3D topologically-connected woody structure through to the high order branching. These QSMs are demonstrated in figure 1.4, where models have been constructed for each of corresponding tree-level point clouds presented in figure 1.3.

Figure 1.3: Individual tree-level point clouds from across a 0.25 ha *Eucalyptus* spp. open forest plot in Karawatha Forest Park, Australia; each uniquely coloured stem has been extracted using the *treeseg* algorithm.

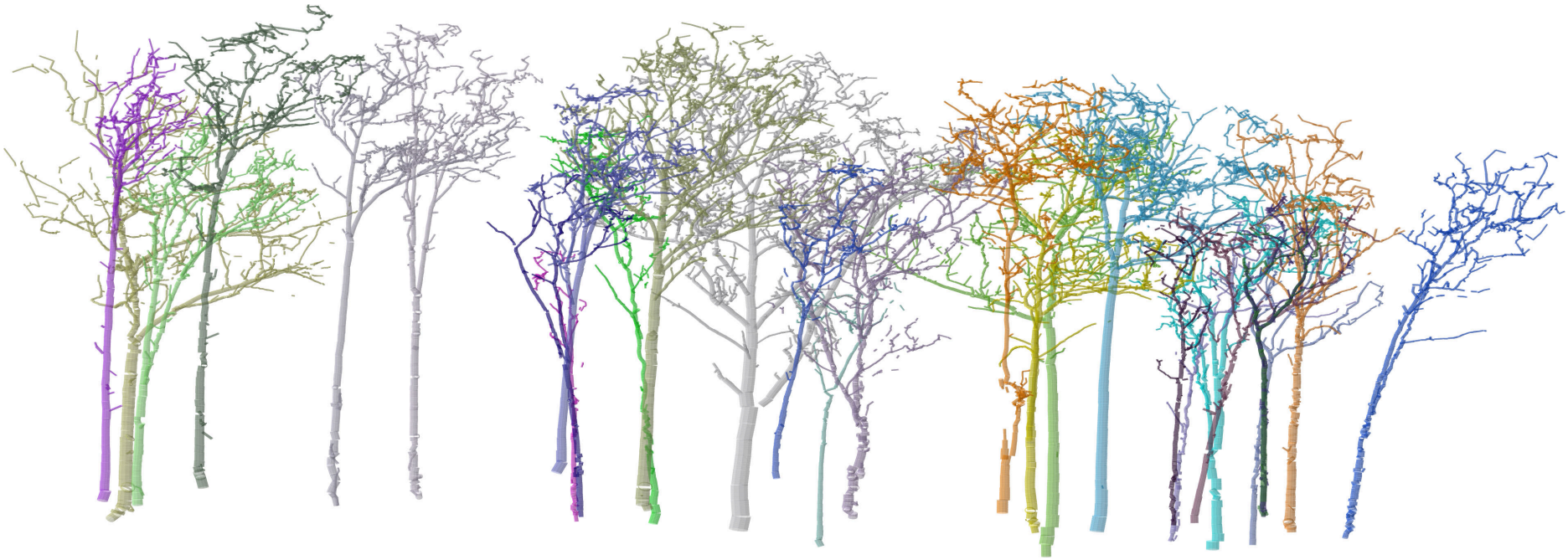


Figure 1.4: Quantitative structural models of the tree-level point clouds shown in figure 1.3, constructed using the Raumonen et al. (2013) shape-fitting algorithm; coloured by individual stem.

The explicit description of 3D tree-level structure afforded by these QSMs permits estimation of tree-scale volume, that alongside an estimate of wood density, allows an estimate of tree-scale AGB to be retrieved. These TLS-derived methods then, provide a non-destructive alternative to allometric estimates of tree-scale AGB, but unlike their allometric counterparts, are not an out-of-sample prediction, rather they are generated from many measurements of the tree in question.

To date, one of the key limitations of these QSM methods has been the inability to deploy them on a wide-scale due to a need for manual user input. Knowledge gap ii), the automatic construction of these QSMs, has been bridged in this thesis through the development of a novel optimisation routine that enables the automatic reconstruction of tree-level point clouds from the Raumonen et al. (2013) shape-fitting algorithm.

Assemblage of these methods then, enables, for the first time, the near-automatic retrieval of tree- and plot-scale AGB from TLS data on a wide-scale. The final knowledge gap is to ascertain that these new methods are indeed capable of accurately estimating tree- and plot-scale AGB from TLS data, beyond the accuracies of that permitted by allometrics. This demonstration has been performed in this thesis through application of these new methods to TLS data acquired from 65 *Eucalyptus* spp. specimens and 3 *Quercus* spp. specimens that have been destructively harvested post-scanning to retrieve a direct measurement of their tree-scale AGB. These direct measurements allow explicit quantification of error in these new TLS-derived estimates.

To demonstrate the new methods developed in this thesis are capable of routinely and near-automatically retrieving tree- and plot-scale AGB from TLS data, they have been applied to TLS data collected during the work for this thesis from 3.25 ha of predominately tropical forest. From these data, 441 individual trees are extracted and reconstructed using the *treeseg* algorithm, the Raumonen et al. (2013) shape-fitting algorithm and the QSM optimisation routine, to retrieve estimates of tree- and plot-scale AGB. These estimates are then directly compared to allometric-derived estimates. A complementary study is also then presented in this thesis with the objective of reducing the uncertainty associated with allometric-derived estimates of AGB by supplementing allometric datasets with these additional TLS data.

1.5 Thesis outline

The primary objective of thesis is:

> **To demonstrate that tree- and plot-scale above-ground biomass can be accurately retrieved on a wide-scale from new 3D measurements of forest structure captured through terrestrial laser scanning.**

The thesis is laid out as follows:

Chapter 2 critiques the current benchmark method for estimating tree- and plot-scale above-ground biomass (AGB), allometry. This chapter presents a novel experimental study demonstrating that non-conformity of current pan-tropical allometric models to underlying statistical assumptions renders statistical inference invalid, and predictive inference biased. Alternatively, in this chapter a new non-linear model is proposed with uncertainty quantification undertaken through non-parametric methods. It is shown that uncertainties in this model exceed 75 % and 25 % of AGB at the tree- and plot-scale respectively, leading to AGB estimates of large swathes of tropical forest being statistically indistinguishable from one another.

Chapter 3 introduces terrestrial laser scanning (TLS), the alternative method considered in this thesis for estimating tree- and plot-scale AGB. This is followed by a review of the literature that has employed TLS data for forest structure retrieval, including the small body of existing work that looked to estimate tree-scale AGB from such data. This chapter concludes by identifying the knowledge gaps that require bridging before it is possible to employ these TLS-derived methods on a wide-scale, and for TLS to be considered a viable alternative to allometrics.

Chapter 4 describes the TLS data that have been collected across 3.25 ha of predominately tropical forest during the work for this thesis in the forests of Australia, Gabon, the United Kingdom, Brazil and French Guiana. This chapter also provides an overview of the instrument used to collect these data, the RIEGL VZ-400.

Chapter 5 introduces the *treeseg* algorithm, developed during this thesis to automatically extract tree-level point clouds from plot-level point clouds. This chapter demonstrates the algorithm by applying it to the data described in chapter 4 to extract 441 tree-level point clouds.

Chapter 6 presents a review of the Raumonen et al. (2013) shape-fitting algorithm that has been employed in this thesis for estimating tree-scale AGB from tree-level point clouds via quantitative structural modelling (QSM). A novel routine is then presented in this chapter to automatically optimise QSMs. This chapter concludes by quantifying the error in tree- and plot-scale estimates of AGB that are derived through these new TLS methods by applying them to validation data acquired across 50 *Eucalyptus* spp. individuals and 3 *Quercus* spp. individuals. It is shown that across these data, the error in estimates of tree- and plot-scale AGB were 23.3 % and 7.9 % respectively

Chapter 7 demonstrates these new TLS-derived methods by applying them to the TLS data described in chapter 4. Tree-scale AGB is retrieved from 441 individuals, and plot-scale AGB is estimated for four forest plots. These results are then compared against allometric estimates, where it is shown that TLS-derived estimates of AGB are 17.9 % larger than their allometric counterparts. This potentially significant finding indicates that allometric estimates of pan-tropical tree- and plot-scale AGB may underestimate global tropical forest carbon storage. A complementary study is then presented in this chapter that aims to reduce allometric uncertainties by augmenting allometric datasets with these new TLS data. It is shown that despite these additional data, uncertainty in allometric-derived estimates of tree- and plot-scale AGB remained largely unchanged.

Chapter 8 presents the conclusions of this thesis and suggestions for further work.

2 Estimates of above-ground biomass from allometry

2.1 Introduction

Chapter 1 section 1.1 introduced the role of the world's forests in sustaining life on earth. Chapter 1 section 1.2 then described the earth observation efforts monitoring the state and change of the above-ground biomass (AGB) of the world's forests. It was shown that all these efforts require calibration from networks of field sites, at which plot-scale AGB has been characterised. Chapter 1 section 1.3 then introduced allometry, the ubiquitous method employed to estimate plot-scale AGB at these sites in lieu of impractical direct measurement of plot-scale AGB via destructive harvest. These empirical allometric relationships allow the inference of tree-scale AGB from simple measurements of tree structure.

The primary objective of this thesis is to demonstrate that new 3D measurements of forest structure from terrestrial laser scanning can provide accurate non-destructive estimates of tree- and plot-scale AGB. To achieve this, it is necessary to benchmark any new TLS-derived estimates, and their corresponding uncertainty, against their 'current best estimate' allometric-derived counterparts. This chapter presents a novel experimental study critiquing pan-tropical allometric models to derive reliable estimates of uncertainty in allometric-derived predictions of tree and plot-scale AGB.

2.1.1 The Chave et al. (2014) allometric dataset and model

This chapter is predicated on the Chave et al. (2014) allometric dataset, that at the time of writing, provides the most comprehensive extant pan-tropical data available. This dataset, a compilation of 43 scientific studies from 53 pan-tropical field sites, comprises a total of 4004 destructively harvested individuals whose AGB (kg) has been measured concurrent with the allometric regressors: bole diameter, D (m), tree height, H (m), and wood density, ρ ($\mathrm{kg\,m^{-3}}$). Figure 2.1 presents the geographical location of these 53 field sites, covering a broad range of tropical, sub-tropical and woodland savannah forest types, in both old-growth and secondary sucesssional types.

The motivation behind these 43 studies was for the advancement of the scientific understanding of forests, with the earliest of these studies undertaken during the 1960s

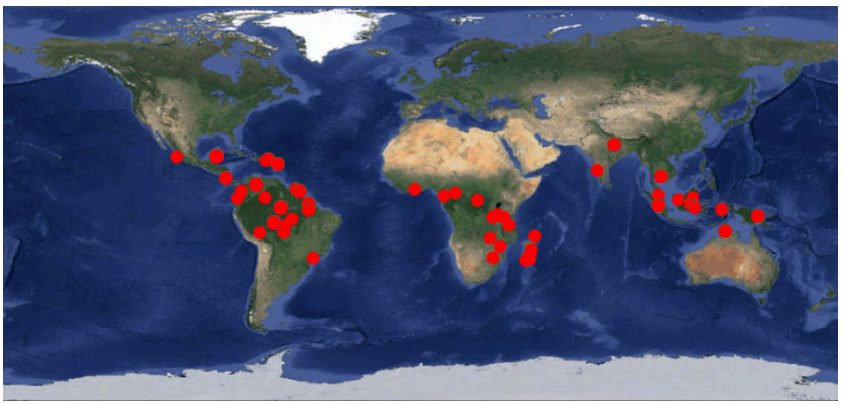


Figure 2.1: The location of the 53 field sites from which the Chave et al. (2014) allometric dataset, comprising 4004 destructively harvested individuals has been constructed (Source: Chave et al. 2014).

and 1970s focusing on ecophysiology (Hozumi et al. 1969; Edwards et al. 1977). From about the turn of the century onwards, in the context of an increasing awareness of global climate change, the focus of these studies was towards efficient carbon accounting (Brown 1997; Ketterings et al. 2001). That is, this allometric dataset is in contrast to those of the first half of the twentieth century that were built by foresters, who aimed to accurately estimate stand-scale merchantable timber (Lovell et al. 2016), where stem volume, as oppose to total tree-scale AGB (i.e., stem and crown woody mass, and leaf mass), were the primary consideration.

To illustrate the Chave et al. (2014) allometric dataset, the distribution of the 4004 stems by D-class is presented in figure 2.2. Across the dataset, D ranges from 0.05–2.12 m, with a mean of 0.24 m, where the number of stems with $D < 0.2$ m contribute 62 % of the total data. Also presented in figure 2.3 is the distribution of the dataset by H-class. The value of H in the dataset ranges from 1.2–70.7 m, with a mean of 16.0 m, where the number of stems with $H < 20$ m contribute 73 % of the total data.

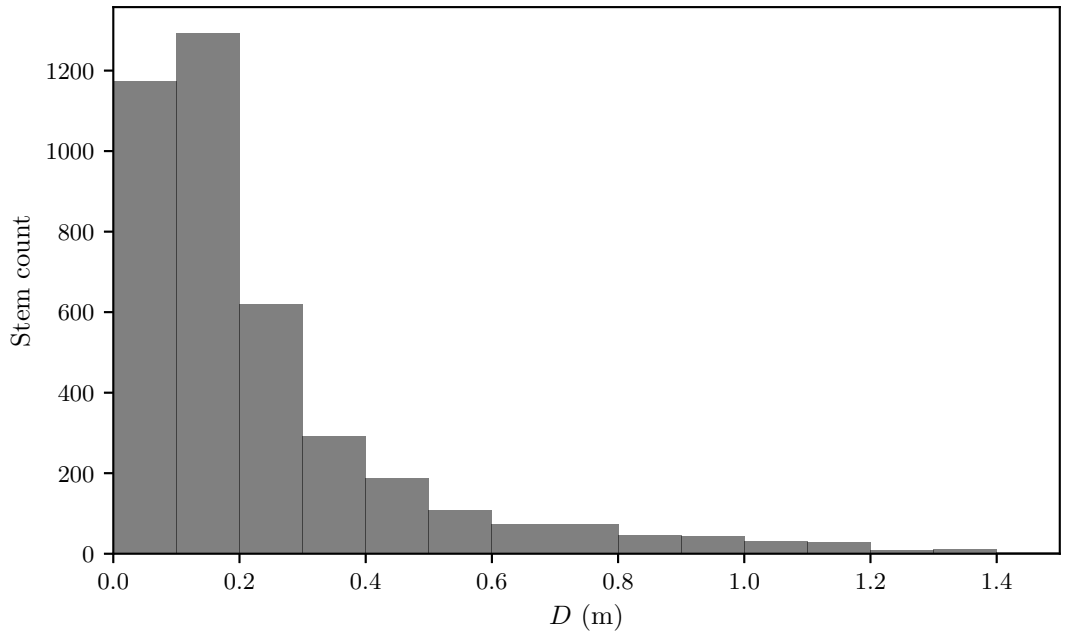


Figure 2.2: The distribution of D across the 4004 individuals in the Chave et al. (2014) allometric dataset.

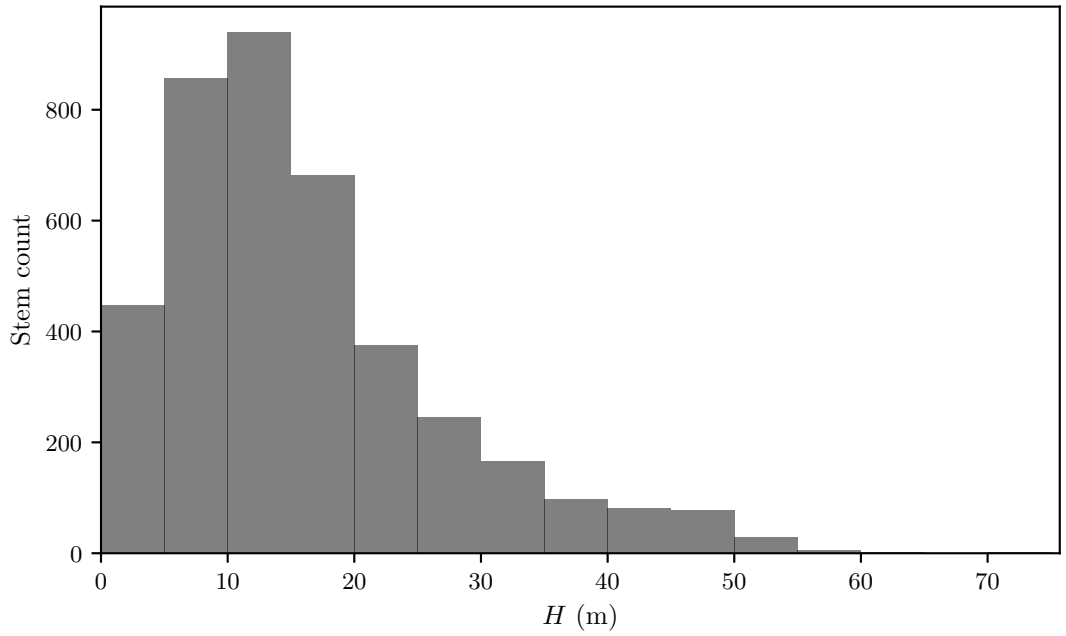


Figure 2.3: The distribution of H across the 4004 individuals in the Chave et al. (2014) allometric dataset.

Figure 2.4 then illustrates the fundamental relationship existing between the structural measurements of each tree in the dataset, and the corresponding AGB, from which the allometric model is derived.

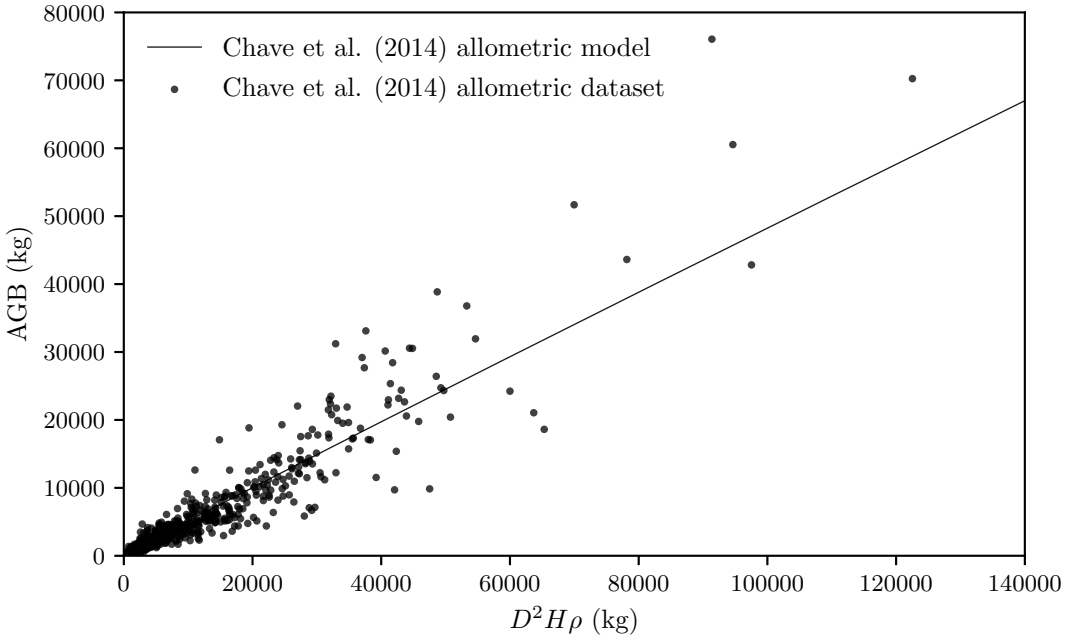


Figure 2.4: The Chave et al. (2014) allometric model overlaid on the allometric dataset.

All the widely used pan-tropical allometric models (Brown 1997; Chave et al. 2005; Feldpausch et al. 2012; Chave et al. 2014) capture this correlation through a single statistical method, namely log-transformed ordinary least squares linear regression. In the instance of the Chave et al. (2014) allometric model, this takes the form:

$$\text{AGB} = 0.636(D^2H\rho)^{0.976} \tag{2.1}$$

This model is illustrated in figure 2.4, overlaid on the allometric dataset,

2.1.2 Uncertainty in allometric-derived estimates of AGB

The importance of quantifying uncertainty in allometric-derived estimates of tree- and plot-scale AGB was described in chapter 1 section 1.3. That is, all up-scaled AGB products, and their subsequent derivatives (e.g., carbon stocks and fluxes), are reliant on calibration from allometric estimates of plot-scale AGB because the direct measurement via destructive harvest is unworkable.

Uncertainty attributed with allometric-derived estimates of AGB can can be broken

into four distinct components. First, error in the measurement of the 4004 individuals that constitute the allometric dataset will induce some error in the resultant allometric model. In the instance of a 'dataset of datasets', such as the one considered here, this could arise from inconsistent measurement protocol and instrumentation.

Second, termed in this thesis as the ecological variance uncertainty, this describes the extent to which population-scale variability is captured by the harvested dataset of 4004 individuals. The magnitude of this term is influenced by considerations ranging from tropical forest species richness to cross-scale spatial variability and environmental factors (Baker et al. 2004; Slik et al. 2015). And, perhaps most importantly, this ecological variance uncertainty is a function of how well the allometric dataset represents large canopy trees that dominate tropical forest AGB distributions (Slik et al. 2013; Lim et al. 2003).

The third source of uncertainty is allometric model uncertainty, arising from the underlying statistical methods that were employed in the construction of the allometric model, and consequentially, how capable the model is of representing the dataset. The final source of uncertainty, measurement uncertainty, arises from permeation of errors in the measurement of the allometric regressors (i.e., D, H and ρ) through the model.

The foremost approach to quantifying uncertainty in estimates of AGB from pan-tropical allometric models is that of Chave et al. (2004). When these methods were applied to the pan-topical allometric model of Chave et al. (2014), the resulting estimate of relative uncertainty in prediction of AGB was approximately 55 % and between 5–10 % at the tree- and 1 ha plot-scale respectively.

This chapter presents a novel experimental study assessing both the validity of log-transformed ordinary least squares linear regression in pan-tropical allometry, and the veracity of the current expectations of uncertainty in allometric-derived estimates of pan-tropical tree and plot-scale AGB.

2.2 Methods

2.2.1 Uncertainty in allometric prediction of tree-scale AGB

As described in the previous section, and when the allometric dataset is considered error free, then relative uncertainty in out-of-sample allometric forecasting of tree-scale AGB, U_t, comprises three additive components (Ahmed et al. 2013). First, tree-scale

allometric model uncertainty, U_{t_A}, describes uncertainty arising from the ability of the model to represent the allometric dataset it was built on. Second, tree-scale measurement uncertainty, U_{t_M}, is introduced during the collection of the allometric predictors of AGB (i.e., D, H and ρ). Third, termed here as tree-scale ecological variance uncertainty, U_{t_E}, this expresses population-scale variability that has not been captured by the allometric dataset.

2.2.1.1 Tree-scale allometric model uncertainty

Log-transformed ordinary least squares regression

The most widely used pan-tropical allometric models are constructed using linear regression, comprising a vector of regressands, $\boldsymbol{y}$, a design matrix of regressors, $\boldsymbol{X}$, and vectors of population parameters and error terms, $\boldsymbol{\beta}$ and $\boldsymbol{\varepsilon}$, respectively:

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \tag{2.2}$$

The ordinary least squares (OLS) solution to estimating $\hat{\boldsymbol{\beta}}$ is the minimisation of the sum of the squared residuals:

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{y} \tag{2.3}$$

To satisfy the Gauss-Markov theorem identifying the OLS solution of $\hat{\boldsymbol{\beta}}$ as the best linear unbiased estimator (BLUE), compliance with the following six assumptions is required (Hayashi 2000): 1) linearity, such that there exists some linear relationship between the regressors and regressand, 2) strict exogeneity, principally meaning that $\boldsymbol{X}$ solely causes $\boldsymbol{Y}$, and 3) the absence of perfect multicollinearity (a full correlation between regressors that would prohibit the inversion of $\boldsymbol{X}$). Estimates of $\hat{\boldsymbol{\beta}}$ remain unbiased but inefficient (i.e., no longer BLUE) in violation of the expectations:

$$E(\varepsilon_i\varepsilon_j|\boldsymbol{X}) = 0 \tag{2.4}$$

$$E(\varepsilon_i^2|\boldsymbol{X}) = \sigma^2 \tag{2.5}$$

4) autocorrelation, any non-random correlation between independent observations of the regressors, and 5) homoscedasticity, such that the error term variance is constant. The

standard error of the regression, σ, that permits statistical inference to be drawn via classical prediction and confidence intervals, remains valid provided assumptions 4 and 5 are met. In addition, these intervals are considered legitimate when 6), the error terms are normally distributed as:

$$\boldsymbol{\varepsilon} \sim \mathcal{N}(0, \sigma^2 \mathbf{I}) \tag{2.6}$$

Here, initially, the validity of pan-tropical AGB inference via log-transformed linear allometric models is tested by assessing the conformity of the following two models, constructed on the Chave et al. (2014) allometric dataset, to the six aforementioned OLS assumptions:

$$\text{M1} : \ln(\text{AGB}) = \beta_0 + \beta_1 \ln(D^2 H \rho) \tag{2.7}$$

$$\text{M2} : \ln(\text{AGB}) = \beta_0 + \beta_1 \ln(D) + \beta_2 [\ln(D)]^2 + \beta_3 [\ln(D)]^3 + \beta_4 \ln(H) + \beta_5 \ln(\rho) \tag{2.8}$$

Regressor amalgamation into the single covariate log-transformed linear model M1 follows the approach of the most widely used pan-tropical allometric models (Brown 1997; Chave et al. 2005; Feldpausch et al. 2012; Chave et al. 2014). Additionally considered here is the log-transformed linear model M2, that allows a non-linear AGB response to D, alongside non-fixed exponents of the regressors (Picard et al. 2012). The conformity of models M1 and M2 to the aforementioned six underlying OLS assumptions is assessed through the following:

1. **Linearity:** The Rainbow statistical test (Utts 1982) is used to evaluate the applicability of the log-transformation of regressor and regressand in both models to attain linearity. Notably, a by-product of the log-transformation is that subsequent AGB prediction requires re-transformation back to arithmetic-space, from which it must be noted that the resulting exponential of equation 2.2 no longer yields the error distribution of equation 2.6. This induced bias is countered in the most widely used pan-tropical allometric models through application of a correction term, formed using σ, as (Neyman et al. 1960; Baskerville 1972):

$$e^{\hat{y}} = e^{\boldsymbol{x}_0 \hat{\boldsymbol{\beta}} + \varepsilon} e^{\frac{\sigma^2}{2}} \tag{2.9}$$

A fundamental implication in employing this correction term is that transformed $\hat{\boldsymbol{\beta}}$ is now dependant on σ to remain unbiased, making predictive inference valid now when OLS assumptions 1 through 6 are met (previously being only 1–3).

2. **Exogenity:** The intended application of the derived allometric models is for prediction of tree-scale AGB, such that the interest lies with the correlation, and not the causation, between the regressand and the regressors. In such cases this permits endogeneity, allowing this assumption to be disregarded here.

3. **Multicollinearity:** A known correlation exists at the population-scale between D and H, that will induce some multicollinearity. A significant increase in multicollinearity is expected in model M2 due to the inclusion of multiple exponent D terms. The magnitude of multicollinearity is gauged through the condition number statistical test (Belsley et al. 1980).

4. **Autocorrelation:** The allometric dataset is comprised of single observations that are assumed to be error-free, allowing this assumption to be disregarded here.

5. **Homoscedasticity:** Non-constant error term variance can be readily identified through visual inspection of the error terms plotted against the regressand. In addition to this, error term variance is assessed through the Breusch-Pagan and White statistical tests (Breusch et al. 1979; White 1980).

6. **Normality:** A Quantile-Quantile plot of the error term distribution against the expected normal distribution is a typical method used to illustrate deviations that are indicative of non-normality. Also, alongside the measure of kurtosis, the Anderson-Darling, Omnibus and Jarque-Bera statistical tests are used to assess the underlying distribution of the error terms (Anderson et al. 1954; D'Agostino 1971; Jarque et al. 1980).

In the results section it is demonstrated that these tests identify both of the considered log-transformed linear models violate the assumptions of homoscedasticity and normality. This results in the return of a spurious estimate of σ, that prohibits statistical inference to be drawn from either model. This is further compounded by the reliance on σ in the correction term that is required for re-transformation, rendering predictive inference biased.

Non-linear least squares regression

Alternatively, a new non-linear least squares, three parameter, single amalgamated regressor model is proposed, taking the form:

$$AGB = \beta_1 (D^2 H\rho)^{\beta_2} + \beta_3 \tag{2.10}$$

This approach is favoured over other linear regression techniques that permit outlier weighting or removal due the assumption that the allometric dataset is error-free (i.e., each data point is considered a rational entry).

The non-linear least squares (NLS) solution to estimation of population parameters, $\hat{\beta}_1$, $\hat{\beta}_2$ and $\hat{\beta}_3$ is the minimisation of the sum of the squared residuals, S:

$$\frac{\partial S}{\partial \hat{\beta}_1} = -2 \sum_{i=1}^{N} (D^2 H\rho)_i^{\hat{\beta}_2} (\mathrm{AGB}_i - \hat{\beta}_1 (D^2 H\rho)_i^{\hat{\beta}_2} - \hat{\beta}_3) = 0 \tag{2.11}$$

$$\frac{\partial S}{\partial \hat{\beta}_2} = -2 \hat{\beta}_1 \hat{\beta}_2 \sum_{i=1}^{N} (D^2 H\rho)_i^{\hat{\beta}_2 - 1} (\mathrm{AGB}_i - \hat{\beta}_1 (D^2 H\rho)_i^{\hat{\beta}_2} - \hat{\beta}_3) = 0 \tag{2.12}$$

$$\frac{\partial S}{\partial \hat{\beta}_3} = -2 \sum_{i=1}^{N} (\mathrm{AGB}_i - \hat{\beta}_1 (D^2 H\rho)_i^{\hat{\beta}_2} - \hat{\beta}_3) = 0 \tag{2.13}$$

With no closed solution, values of $\boldsymbol{\beta}$ are evaluated numerically through the Levenberg-Marquardt iterative linearisation approximation (Marquardt 1963), solving for p, over steps k:

$$(\boldsymbol{J}_k^T \boldsymbol{J}_k + \lambda_k \boldsymbol{D}_k^T \boldsymbol{D}_k) p_k = -\boldsymbol{J}_k^T (\beta_1 (D^2 H\rho)^{\beta_2} + \beta_3)_k \tag{2.14}$$

Until convergence criteria are satisfied, with the diagonal, $\boldsymbol{D}$, of the Jacobian, $\boldsymbol{J}$, described by the residuals, ε as:

$$\boldsymbol{J} = \begin{bmatrix} \frac{\partial \varepsilon_1}{\partial \beta_1} & \frac{\partial \varepsilon_1}{\partial \beta_2} & \frac{\partial \varepsilon_1}{\partial \beta_3} \\ \vdots & \vdots & \vdots \\ \frac{\partial \varepsilon_N}{\partial \beta_1} & \frac{\partial \varepsilon_N}{\partial \beta_2} & \frac{\partial \varepsilon_N}{\partial \beta_3} \end{bmatrix} \tag{2.15}$$

Assumptions underpinning NLS are nominally accordant with OLS assumptions 1-6 (Graybill et al. 1994), where linearity is replaced with the new non-linear functional form. Here, the expectation is for the resultant non-linear model constructed on the

Chave et al. (2014) allometric dataset to violate assumptions 5 and 6, exhibiting strong heteroscedasticity and non-normality that far exceeds that of either log-transformed linear model. This expectation is at the heart of the debate on the appropriateness of non-linear modelling in allometrics (Packard et al. 2011; Mascaro et al. 2011; Xiao et al. 2011).

Despite this, because the non-linear model has been constructed in arithmetic-space, non-conformity by the model to these two assumptions means only that $\hat{\boldsymbol{\beta}}$ are no longer asymptotically efficient; critically they will remain weakly consistent, and thus unbiased (Greene 2011). This is in contrast to either log-transformed linear model in the presence of heteroscedasticity or non-normality because of the requirement for later re-transformation. σ however, would, analogous to the log-transformed linear models, be returned biased from a poorly specified error term variance-covariance matrix; implying that statistical inference from the non-linear model must be drawn through non-parametric methods.

To infer U_{t_A} from the non-linear model, prediction intervals are utilised. Values are quantified through the non-parametric method of non-linear quantile regression (Koenker 2005), with the non-linear model for the τth quantile described as:

$$Q_{\mathrm{AGB}_i}(\tau|(D^2H\rho)) = g((D^2H\rho), \hat{\boldsymbol{\beta}}(\tau)) \tag{2.16}$$

Where g defines the quantile response function; $\hat{\boldsymbol{\beta}}(\tau)$ is estimated through the minimisation:

$$\hat{\boldsymbol{\beta}}(\tau) = \min_{\hat{\boldsymbol{\beta}} \in \Re} \sum_{i=1}^{N} \rho_\tau(\mathrm{AGB}_i - g((D^2H\rho)_i, \hat{\boldsymbol{\beta}}(\tau)) \tag{2.17}$$

Application of the interior point algorithm described in Koenker et al. (1996) provides a numerical solution to equation 2.17. U_{t_A} is then defined from the upper and lower quantiles, τ_u and τ_l, as:

$$U_{t_A} = \frac{1}{2}[(\beta_1(\tau_u)(D^2H\rho)^{\beta_2(\tau_u)} + \beta_3(\tau_u)) - (\beta_1(\tau_l)(D^2H\rho)^{\beta_2(\tau_l)} + \beta_3(\tau_l))]\mathrm{AGB}^{-1} \tag{2.18}$$

2.2.1.2 Tree-scale measurement uncertainty

Uncertainty that propagates through the allometric estimate of tree-scale AGB that is attributed to error in the measurement of the allometric regressors (i.e., D, H and ρ), U_{t_M}, is typically linearly approximated through a first order Taylor series expansion, yielding, for the non-linear model:

$$\sigma_m = \sqrt{(2\beta_2\frac{\sigma_D^2}{D})^2 + (\beta_2\frac{\sigma_H^2}{H})^2 + (\beta_2\frac{\sigma_\rho^2}{\rho})^2} \tag{2.19}$$

Where the measurement variance is defined as σ_D^2, σ_H^2 and σ_ρ^2.

The approach adopted by both the log-transformed linear model M1 and the non-linear model to amalgamate the regressors ensures $f(\text{AGB})$ is significantly non-linear, dictating that the above exclusion of higher order terms in the Taylor series expansion may lead to an underestimation of σ_m.

A simplistic alternative used here to quantify σ_m is Monte Carlo simulation. In line with Chave et al. (2004), the following measurement variance distributions are assumed:

$$\hat{D} \sim \mathcal{N}(D, 0.05) \tag{2.20}$$

$$\hat{H} \sim \mathcal{N}(H, 0.1) \tag{2.21}$$

$$\hat{\rho} \sim \mathcal{N}(\rho, 0.1) \tag{2.22}$$

N random draws are made from these distributions from which the AGB distribution, AGB*, is generated. σ_m is then quantified from this distribution as:

$$\sigma_m = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\text{AGB}_i^* - \text{AGB})^2} \tag{2.23}$$

Under these assumed normal measurement distributions, U_{t_M}, for the standard score, z, at the interval, α, can be defined as:

$$U_{t_M} = \sigma_m z_\alpha \text{AGB}^{-1} \tag{2.24}$$

2.2.2 Uncertainty in allometric prediction of plot-scale AGB

Considered next is the relative uncertainty associated with allometric forecasting of plot-scale AGB, U_p, comprising three additive terms; plot-scale allometric model uncertainty, U_{p_A}, plot-scale measurement uncertainty, U_{p_M} and plot-scale ecological variance uncertainty, U_{p_E}.

2.2.2.1 Plot-scale allometric model uncertainty

To infer U_{p_A} from the non-linear model, confidence intervals are utilised. Due to the expectation of the non-linear model to violate the assumptions of homoscedasticity and normality, confidence intervals are generated through the wild bootstrap (Wu 1986), a non-parametric method particularly well-suited to data exhibiting strong heteroscedasticity.

From the Chave et al. (2014) dataset, F, the random sample, F^*, is drawn, whose AGB is redistributed through perturbation of the residuals, ε, according to the Rademacher distribution:

$$\mathrm{AGB}_i^* = (\hat{\beta}_1 (D^2 H \rho)_i^{\hat{\beta}_2} + \hat{\beta}_3) + \varepsilon_i \upsilon_i \tag{2.25}$$

$$\upsilon_i = \begin{cases} 1 & P(\frac{1}{2}) \\ -1 & P(\frac{1}{2}) \end{cases} \tag{2.26}$$

Population parameters, $\hat{\beta}^*$, estimated for each member of F^*, produce the distribution, $\theta_{\mathrm{AGB}[*]}$. Confidence intervals, for the upper and lower intervals, α_u and α_l respectively, are calculated from interpolated percentiles. U_{p_A}, is then described through the summation of these confidence intervals contributed by the N stems inside the considered plot:

$$U_{p_A} = \frac{1}{2} \sum_{i=1}^{N} [\theta_{\mathrm{AGB}[\alpha_u],i} - \theta_{\mathrm{AGB}[\alpha_l],i}] (\sum_{j=1}^{N} \mathrm{AGB}_j)^{-1} \tag{2.27}$$

2.2.2.2 Plot-scale measurement uncertainty

From the assumed normal measurement probability distributions, averaging over the N stems inside the considered plot and disregarding covariance, U_{p_M} is described as:

$$U_{p_M} = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} 2 z_\alpha \sigma_{m,i} \mathrm{AGB}_j{}^{-1} \tag{2.28}$$

2.3 Results

2.3.1 Tree-scale uncertainty

Log-transformed ordinary least squares regression

Table 2.1 presents the results of the statistical tests used to assess the conformity of the log-transformed linear models M1 and M2 to the following relevant underlying OLS assumptions:

- **Linearity:** The log-transformation of regressor and regressand satisfies linearity, confirmed for both models through the Rainbow statistical test that fails to reject the null hypothesis of linearity.

- **Multicollinearity:** Some multicollinearity is observed in model M1 through the condition number test, whilst significant collinearity is present in model M2. However, neither model exhibit perfect multicollinearity.

- **Homoscedasticity:** Both models accept the alternative hypothesis of heteroscedasticity in the two statistical tests. Figure 2.5 presents the distribution of the error terms across the regressand for model M1, where variable bias can be observed throughout the distribution, alongside reduced tail spread. This is consistent with the statistical tests in indicating heteroscedasticity.

- **Normality:** Figure 2.6 compares the distribution of residuals against the expected normal distribution for model M1. The presence of fat tails and bowing highlights deviation from the normal distribution. Evidence of non-normality is further accrued through leptokurtic kurtosis, and acceptance of the alternative hypothesis of non-normality in the three statistical tests.

Test	M1			M2		
	Statistic	**P-Value**	**Indicative of:**	**Statistic**	**P-Value**	**Indicative of:**
Rainbow	0.978	0.689	Linearity	0.952	0.862	Linearity
Condition	16.2	-	Multicollinearity	344	-	Multicollinearity
Breush-Pagan	36.792	0.000	Heteroscedasticity	15.967	0.000	Heteroscedasticity
White	21.368	0.000	Heteroscedasticity	6.093	0.000	Heteroscedasticity
Kurtosis	3.966	-	Non-normality	4.057	-	Non-normality
Anderson-Darling	5.855	0.000	Non-normality	6.618	0.000	Non-normality
Omnibus	76.528	0.000	Non-normality	93.298	0.000	Non-normality
Jarque-Bera	156.873	0.000	Non-normality	194.313	0.000	Non-normality

Table 2.1: Results from the statistical tests used to assess the conformity of linear models M1 and M2 to the relevant underlying assumptions of OLS modelling.

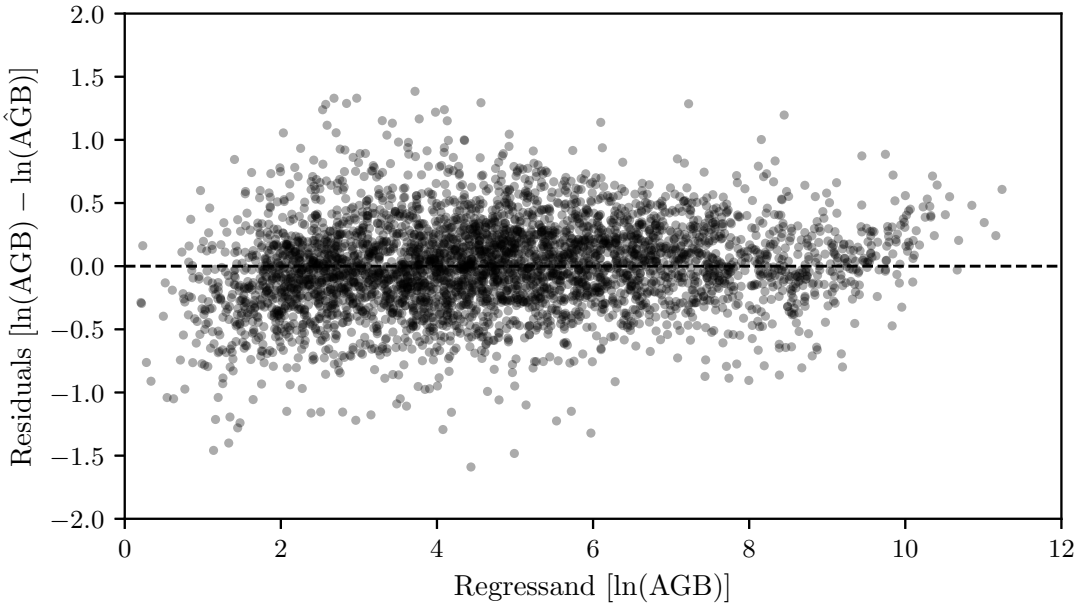


Figure 2.5: Distribution of error terms across the regressand for linear model M1. Reduced tail spread and the variable bias (negative to positive) observed through the distribution, alongside the results of the Breush-Pagan and White statistical tests indicate heteroscedasticity.

The corroborating evidence from the visual inspection of figures 2.5 and 2.6, combined with the results of multiple statistical tests demonstrates that both log-transformed linear models violate the underlying assumptions of homoscedasticity and normality. As such, neither of these models can be used to retrieve valid statistical or predictive inference of pan-tropical AGB.

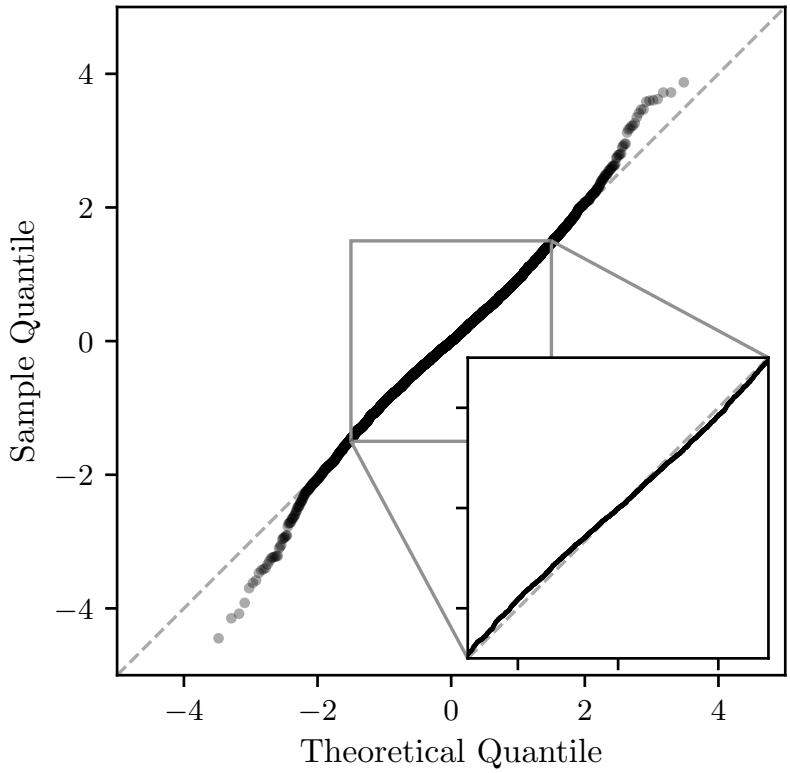


Figure 2.6: Comparison of the linear model M1 error term distribution (sample) against the normal distribution (theoretical). Non-normality is indicated through fat tails and bowing; this is additionally supported by the results of three statistical tests and leptokurtic kurtosis.

Non-linear least squares regression

The proposed non-linear model, generated using the Chave et al. (2014) allometric dataset, takes the form:

$$\text{AGB} = 0.179(D^2 H \rho)^{1.102} + 80.235 \tag{2.29}$$

Figure 2.7 presents this model overlain on the allometric dataset. Additionally presented in the figure are the 5 % and 95 % prediction and confidence intervals, constructed about the model via the described non-parametric uncertainty quantification framework.

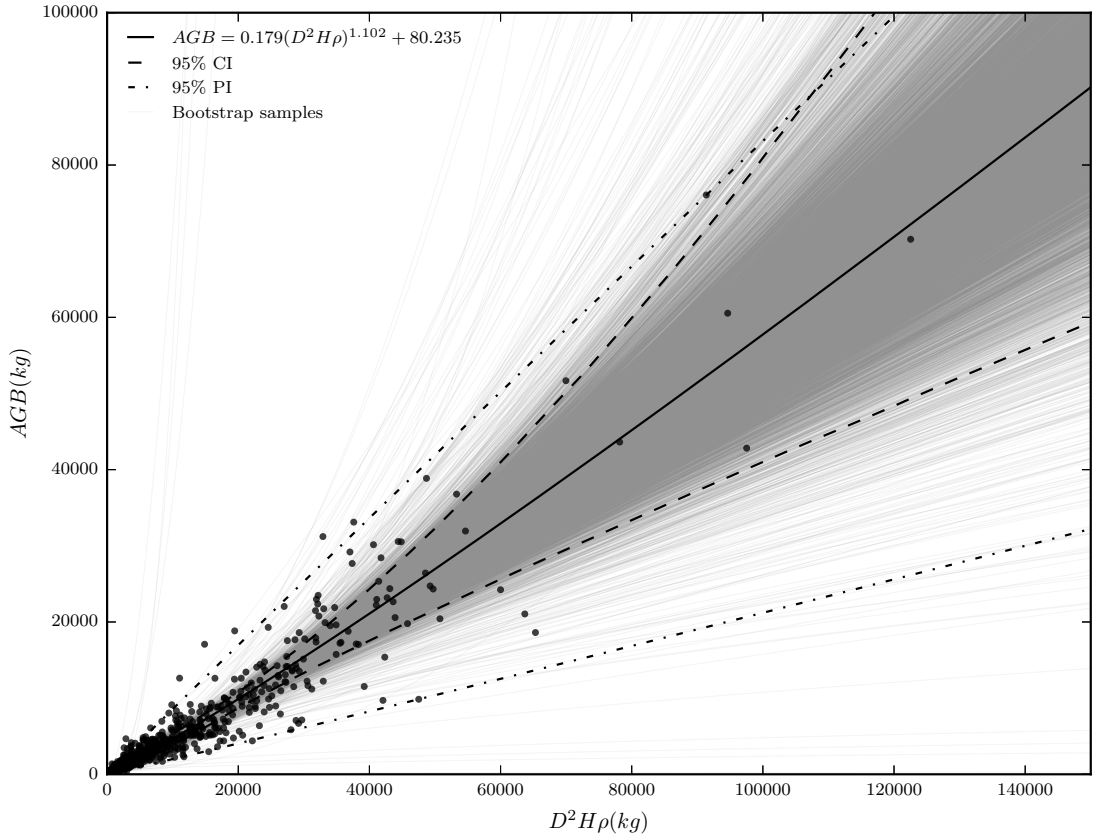


Figure 2.7: The non-linear least squares model overlain on the Chave et al. (2014) allometric dataset. The 95 % prediction intervals, representing tree-scale allometric model uncertainty, are generated through non-linear quantile regression. The 95 % confidence intervals, used to represent plot-scale allometric model uncertainty, are generated through the wild bootstrap.

This model and framework has been applied to data acquired from four 1 ha tropical forest field sites in Gabon and Ghana. These data, drawn from the Global Ecosystem Monitoring network (Marthews et al. 2014), encompass 1666 stems ($D > 0.1$ m) whose

D, H and ρ have been measured.

Per stem, U_t has been characterised through the summation of U_{t_A}, derived via non-linear quantile regression ($\tau = 0.05$) and U_{t_M}, derived via Monte Carlo simulation ($\alpha = 0.05$). Across the 1666 stems, U_t was highly variable, ranging from 18.1–115.2 %, with a mean of 76.7 %. At this mean, U_{t_A} and U_{t_M} contributed 54.5 % and 22.2 % to U_t respectively. The contribution by both of these components to U_t across all the assessed stems with AGB $< 10\,000$ kg is illustrated in figure 2.8. It can be seen that for stems with AGB > 1000 kg, U_t largely plateaus, and can be approximated at 115 %.

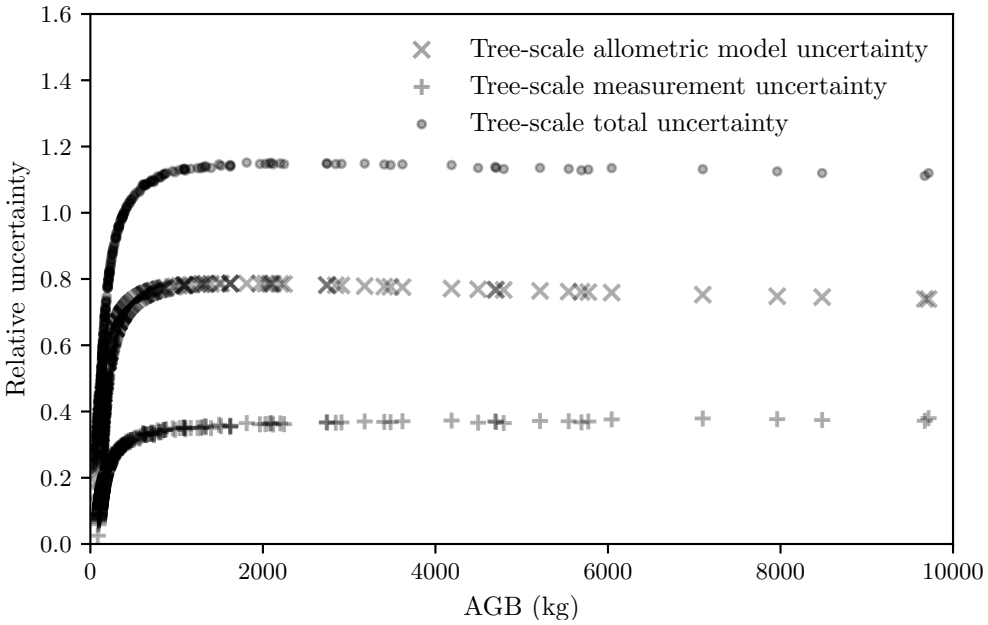


Figure 2.8: Contribution to total tree-scale uncertainty by allometric model and measurement uncertainty across the 1666 individuals inside the four considered 1 ha plots with $AGB < 10\,000$ kg. Mean total tree-scale uncertainty averaged 76.7 %; above AGB > 1000 kg this can be largely approximated at 115 %.

2.3.2 Plot-scale uncertainty

For the four field sites, plot-scale AGB has been quantified from the summation of tree-scale AGB for each stem inside the plot. Per plot, U_p has been characterised through the summation of U_{p_A}, derived via the wild bootstrap ($\alpha = 0.05$) and U_{t_M}, derived via Monte Carlo simulation ($\alpha = 0.05$). The results from each plot are presented in table 2.2. Across the four plots, U_p ranged from 21.7–30.3 % with a mean of 26.1 %. At the mean uncertainty, U_{p_A} and U_{p_M} contributed 24.6 % and 1.5 % respectively.

Plot code	AGB (kg)	AGB_{ll} (kg)	AGB_{ul} (kg)	U_p (%)	$\bar{U}_t$ (%)
ANK-01	300 089	222 050	385 346	27.2	73.9
LPG-01	470 224	347 302	634 450	30.5	77.1
MNG-03	525 984	415 776	636 325	21.0	80.3
MNG-04	379 779	289 652	479 437	25.0	81.3

Table 2.2: Plot-scale estimates of AGB and corresponding uncertainty across the four sets of considered field data.

Plot-scale AGB and U_p for each plot is displayed in figure 2.9. The observed asymmetry in U_p is a function of the plot composition. Across the plots, this averaged such that the upper interval was 10.8 % larger than the lower interval. In one plot, where a single tree contributed 26.3 % of plot-scale AGB, this weighting rose to 24.6 %.

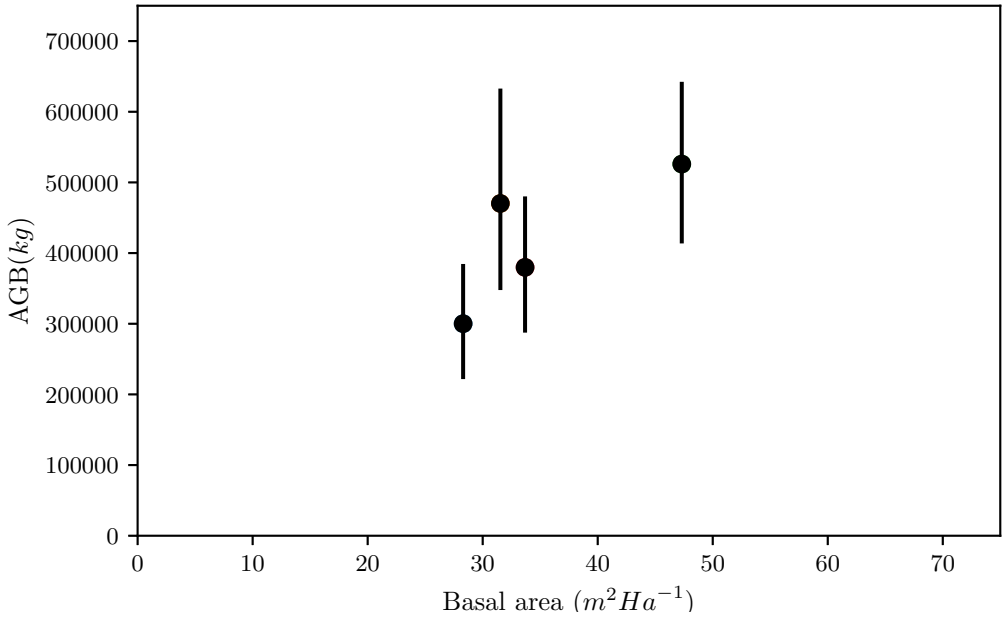


Figure 2.9: Plot-scale AGB and corresponding uncertainty intervals across the four 1 ha field plots. Mean plot-scale uncertainty averaged 26.1 %, whilst across the four plots the upper interval was on average 10.8 % larger than the lower interval.

2.4 Discussion

It has been demonstrated that the two considered pan-tropical log-transformed linear allometric models do not comply with the underlying OLS assumptions of homoscedasticity and normality. Consequentially, this results in spurious estimation of σ, that prohibits the application of classical derivatives including prediction and confidence intervals for statistical inference. Additionally, the dependence of σ in the correction term that is required in the re-transformation of the model to arithmetic-space renders tree-scale es-

timates of AGB biased. From these results, the conclusion is drawn that log-transformed ordinary least squares regression is ill-suited to pan-tropical allometry.

In this chapter, a new non-linear allometric model has been proposed that is built on the same allometric dataset, where uncertainties have been characterised through non-parametric methods. Through application of this new model to real field data, it has been shown that relative uncertainty in estimates of tree-scale AGB averaged approximately 75 %, and regularly exceeded 115 % in the larger stems. Whilst performance significantly improves when up-scaling to the plot, where uncertainty in the estimate of AGB approximated 25 %, this uncertainty is markedly divergent from counterparts in the literature.

The implication of such large uncertainties is that the AGB of large swathes of tropical forest are statistically indistinguishable from one another when inferred through this non-linear pan-tropical allometric model. Of the four plots considered, only the smallest and largest, with an estimated plot-scale AGB of 300 089 kg and 525 983 kg respectively, were statistically discernible from one another. A further implication of these uncertainties is that inferring temporal change via this model will not be feasible for all but the largest of changes or disturbances (i.e., plot-scale AGB must change by approximately 25 % before detection is permissible).

Strikingly, the uncertainty quantification framework presented here omitted a number of sources of uncertainty, leading to the conclusion that this estimate of uncertainty is probably conservative. Prominently, it is unlikely that the measurement variances will adhere to the normal distributions that have been assumed. If these distributions are indeed non-normal, then the observed contraction in the measurement uncertainty when up-scaling from the tree to plot (22.2 % down to 1.5 %), would largely disappear, whilst also inducing some degree of bias. Further uncertainty would also likely arise from measurement covariance, species mis-recognition and the accuracy at which a 1 ha plot can be established.

A further key source of uncertainty omitted in this experimental study is the influence of error in the allometric dataset. As described in section 2.1.2, the Chave et al. (2014) allometric dataset is a 'dataset of datasets', whose underlying data were collected across a time frame spanning six decades. Speculatively, it would be highly unlikely that the precision and accuracy of the destructive harvest measurements would remain constant

across studies, although provided such errors were normally distributed, there would be minimal impact on the correctness of the regression line. However, if a systematic bias was introduced, such as changing the protocol for the measurement of average wood density, then the expectation would be for a shift in the regression line, inducing some further uncertainty in the estimates of tree- and plot-scale AGB.

The final, and perhaps most significant omission from this analysis is what has been termed here the ecological variance uncertainty. This source of uncertainty arises from the allometric dataset not fully capturing population-scale variability (i.e., the variation of the AGB of each tree across the tropics with respect to the allometric regressors). Whilst the most comprehensive allometric dataset available has been used here, given the ad-hoc nature of these data, it is highly unlikely they are of sufficient size, species and geographical range to capture the population-scale variability. Furthermore, with only 90 stems from these data with $D > 0.5\,\mathrm{m}, H > 35\,\mathrm{m}$, the variability in larger stems is likely to be poorly sampled and/or biased. This is an unfortunate misalignment between allometric datasets and field sites, where relative to stem count, the largest stems contribute disproportionally to plot-scale AGB. Here, this mismatch manifests through the non-parametric uncertainty quantification framework as asymmetric prediction and confidence intervals, where those plots with largest basal area and AGB exhibited greatest asymmetry.

However, the large variance exhibited by the 4004 stems in the allometric dataset alone, as observed in figure 2.7, suggests that a lack of harvested data is not the root cause of the large uncertainties associated with this pan-tropical model. This analysis would suggest that a more fundamental limitation here is that the model regressors D, H and ρ are essentially incapable of describing AGB to within any reasonable tolerance. That is, whilst no attempt has been made here to derive causation because of model endogeneity, AGB is driven by an array of environment-, geography-, species- and age-specific considerations that are simply not reducible to three simplistic structural parameters.

2.5 Conclusions

This chapter has presented a novel experimental study critiquing pan-tropical allometrics. It has been determined that the sole statistical method used in the construction of

all the widely used pan-tropical allometric models, log-transformed ordinary least squares linear regression, is inappropriate. This has been demonstrated through application of the Chave et al. (2014) allometric dataset to identify that heteroscedasticity and a non-normal distribution of error terms in common form linear allometric models will return spurious statistical inferences. Furthermore, it has been highlighted that a reliance by these models on the correct specification of the errors terms for re-transformation from log-space, will additionally render predictive inferences biased.

Alternatively, in this chapter, a non-linear least squares model has been proposed that has also been constructed on the Chave et al. (2014) allometric dataset. Chiefly, it has been noted that the model, having been constructed in arithmetic-space, will, in contrast to log-space linear models, provide unbiased estimates of AGB. However, the model is subject to similarly misleading statistical inferences when drawn from classical methods; a problem solved here through an uncertainty quantification framework developed using non-parametric methods.

This new non-linear pan-tropical allometric model has been applied to four field sites, where it was shown that relative uncertainty in estimates of tree- and plot-scale AGB exceeded 75 % and 25 % respectively; a significant divergence from the current expectations in the literature. These large uncertainties result in only the smallest and largest of the four plots, whose AGB was estimated at 300 089 kg and 525 983 kg respectively, being statistically identifiable from one another.

These results demonstrated that the non-linear allometric model proposed here is not fit for purpose because its estimates of AGB across large swathes of tropical forest will be statistically indistinguishable from one another. A subsequent implication of these results is that temporal change detection through this model is permissible in only the most extreme of disturbances.

Due to the reliance of up-scaled AGB products on allometry, these results highlight a need to re-assess the goodness of current pan-tropical AGB products across all scales. These results also suggest that there is a requirement for the introduction of alternative techniques that are capable of more accurately estimating tree- and plot-scale AGB. Chapter 3 provides a literature review of the alternate method considered in this thesis, terrestrial laser scanning.

3 Forest structure from terrestrial laser scanning

Chapter 2 presented a novel experimental study that showed uncertainty in estimates of above-ground biomass (AGB) from pan-tropical allometric models exceeds 75 % and 25 % of AGB at the tree- and plot-scale respectively. These results demonstrated that there is a requirement for the introduction of new methods that are capable of more accurately estimating tree- and plot-scale AGB.

This chapter provides a review of the alternative method considered in this thesis, terrestrial laser scanning (TLS). Section 3.1 presents an overview of the fundamental principles of light detection and ranging (LiDAR) more generally, and then TLS specifically. This begins in section 3.1.1 with an introduction to the general LiDAR equation, followed by section 3.1.2 describing ranging estimation and waveform decomposition. Section 3.1.3 reviews the radiometric calibration of TLS instruments. Sections 3.1.4 and 3.1.5 then introduce the role of co-registration and protocols for TLS data acquisition respectively. Finally, section 3.1.5 describes the current state of commercial and research TLS instruments.

Section 3.2 then reviews the body of published literature that looked to retrieve parameters of forest structure from TLS data. This is broken down into section 3.2.1 that briefly reviews work focused on plot-scale retrieval of forest structure. Section 3.2.2 then evaluates the literature concentrating on the tree-scale, and in particular, the estimation of AGB.

Finally, section 3.3 concludes this chapter by stating the knowledge gaps that currently prohibit the retrieval of tree- and plot-scale AGB from TLS data on a wide-scale (here meaning deployment across tens of 1 ha forest plots).

3.1 Terrestrial laser scanning

Laser range finding dates back to introduction by various military in the early 1960s (Nambiar 2004). In the most simple form, a pulse of radiation is emitted from some instrument towards a target object, and upon interception with the target, will scatter, with some proportion returning to the detector of the instrument. Assuming a single scattering event, and that the emitter and detector are coincident, the range from in-

strument to target, R, can be calculated from the time-of-flight of the pulse, t, and the speed of light, c, as:

$$R = \frac{tc}{2} \tag{3.1}$$

This measurement of R can be converted into Euclidean space (i.e., 3D space), relative to the optical centre of the instrument, after assuming some convention on azimuth, θ, and zenith, φ, as:

$$x = R\sin\theta\cos\varphi \qquad y = R\sin\theta\sin\varphi \qquad z = R\cos\theta \tag{3.2}$$

Many such measurements made across the extent of a scene will provide a 3D sampling of the scene, realised through the construction of a point cloud. An example is shown in figure 3.1 of a point cloud generated from over 20 million pulses fired through a tropical forest in Lopé National Park, Gabon.

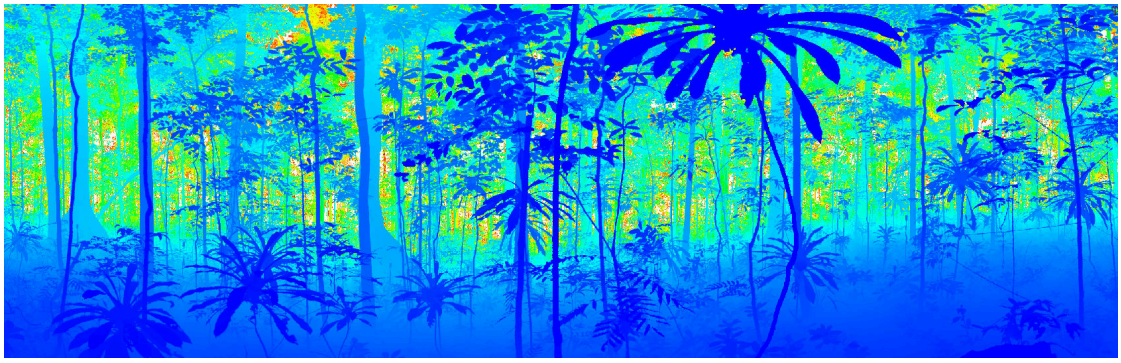


Figure 3.1: Panoramic view of a point cloud acquired in tropical forest in Lopé National Park, Gabon; coloured by range from instrument to target.

3.1.1 The general LiDAR equation

In the scenario described above, the power of the radiation returning to the detector, P_R, is described, as a function of the power of the radiation emitted, P_E, through the general LiDAR equation (Jelalian 1992; Pfeifer et al. 2008):

$$P_R = P_E \underbrace{[\frac{4}{\pi\Theta^2 R^2}]}_{1} \underbrace{[\frac{\rho A}{\Omega R^2}]}_{2} \underbrace{[\frac{\pi D^2}{4}]}_{3} \underbrace{[\eta_A \eta_S + P_B]}_{4} \tag{3.3}$$

With the four denoted terms of attenuation defined separately and in turn below.

3.1.1.1 Beam spread

The first term of equation 3.3 describes the attenuation of P_E due to the divergence of the laser beam, Θ, with respect to the inverse square of R. All laser beams experience far field divergence (i.e., after the focal point, F) due to diffraction. Θ is an important instrument characteristic, as the diameter of the beam footprint at the target, D_R, alongside the accuracy in the estimation of R, determine the smallest unambiguous distance that is resolvable in a point cloud. D_R can be described as a function of R, Θ, and the diameter of the beam footprint as it exits the instrument (a small distance after F), D_E, as:

$$D_R = D_E + R \tan \Theta \tag{3.4}$$

3.1.1.2 Backscatter cross-section

The second term of equation 3.3 characterises how the properties of the target influence the attenuation of P_E. The target reflectance, ρ, defines the proportion of the intercepted radiation that reflects from the surface of the target (i.e., not absorbed or transmitted) in lossless and diffuse conditions (Schaepman-Strub et al. 2006). The spectral reflectance (i.e., as a function of the radiation wavelength) of some relevant materials is presented in figure 3.2.

Next, Ω, denotes the directionality of the radiation that is reflected from the surface of the target, which is primarily a function of the angle of incidence between the beam and the assumed non-Lambertian target (Kaasalainen et al. 2011). Finally, A, defines the proportion of the beam footprint that is intercepted by the target, assuming a constant irradiance across the footprint.

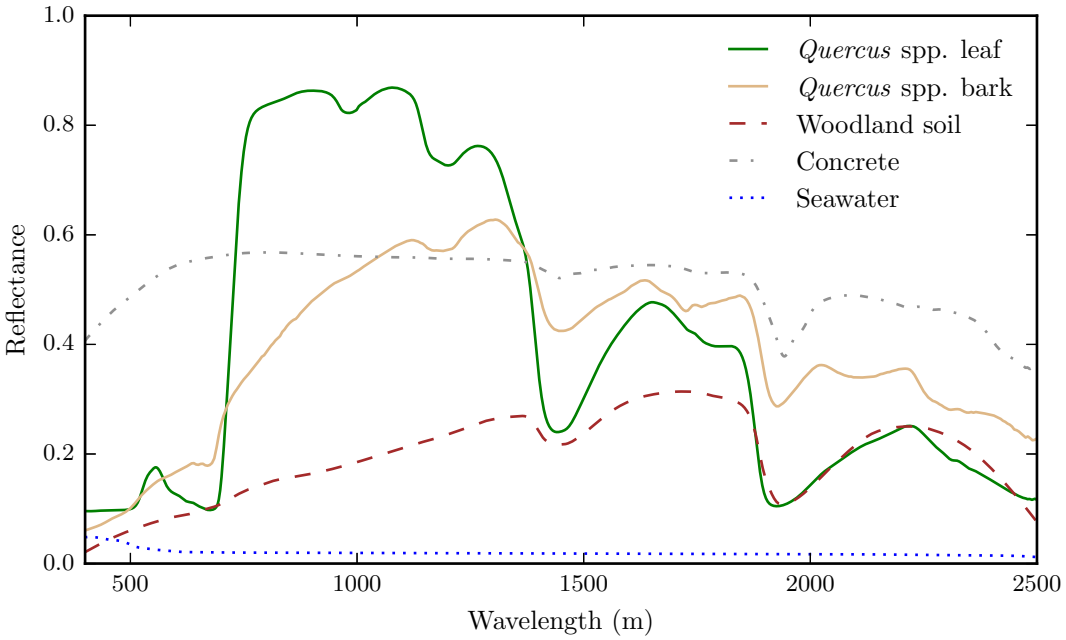


Figure 3.2: Spectral reflectance of *Quercus* spp. leaf and bark alongside woodland soil, concrete and seawater (Data source: Clark et al. (2007))

3.1.1.3 Received power

The third term of equation 3.3 describes the proportion of radiation scattered towards the instrument that is intercepted by the detector (Wagner et al. 2006). D, denotes the diameter of the aperture of the detector.

3.1.1.4 Transmission

The fourth and final term of equation 3.3 describes the attenuation of P_E as the pulse transits through both the atmosphere, η_A, and the instrument η_S. η_A is primarily driven by humidity and precipitation causing attenuation through absorption and non-selective scattering respectively.

η_S is determined by the instrument components, with potential sources of inefficiency arising from, but not limited to, the lenses, filters and mirrors. Finally, the additive term, P_B, denotes the power of any background radiation at the same wavelength.

3.1.2 Ranging and waveform decomposition

As defined in equation 3.1, R is determined by the round-trip time of the pulse from emitter to detector. In practice, the detector will comprise one or more photodiodes (where incoming radiation induces an electric current) that discretely sample radiation

at the emitted wavelength. For some TLS instruments, usually dubbed 'first return instruments', R will be estimated from the timestamp of the discrete sample at which the electric current induced by P_R exceeds the signal-to-noise ratio. The accuracy of the estimate of R then, is directly related to the time between these discrete samples, t_s, and the rise time of the laser pulse.

Other more advanced TLS instruments, often referred to as 'waveform instruments', will continuously collect and store these discrete samples. Figure 3.3a illustrates one such raw signal (Shan et al. 2008). In this instance, R is estimated by decomposing the waveform and estimating the position of peak amplitude through modelling approaches. A number of approaches exist for waveform decomposition (Hancock et al. 2015), in the example of figure 3.3b, through a non-linear least squares fit numerically solved by the Levenberg-Marquardt algorithm.

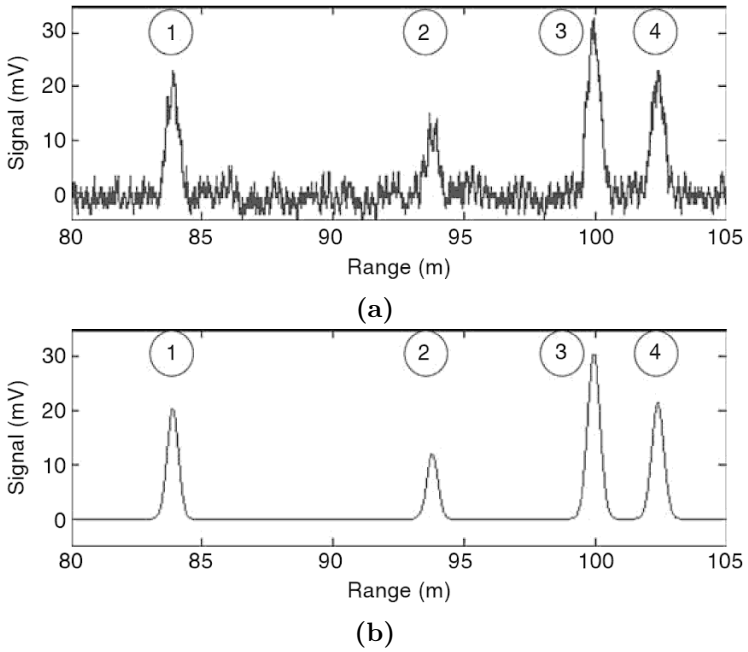


Figure 3.3: (a) Linearly interpolated discrete samples of a returning pulse, and (b) the modelled waveform using non-linear least squares; in this instance four targets have been digitised from the waveform (Source: Shan et al. 2008).

Waveform instruments have three primary benefits over their simpler first return counterparts. First, the accuracy and precision in the estimation of R is usually higher. This is because the modelled waveform is used to infer between the discrete samples.

Second, in the situation where $A \neq 1$ (i.e., the footprint is not fully occupied by the target), the remaining proportion of the footprint not intercepted by the first target may potentially be intercepted by a second target. If enough radiation from this second

scattering event returns to the detector, and is recorded in the waveform, then the range to both targets can be resolved. This is demonstrated in figure 3.3, whereby four targets have been digitised from a single pulse.

3.1.3 Radiometric calibration

The third benefit of waveform instruments is the potential to derive additional parameters of the target other than R. When considering the general LiDAR equation in equation 3.3, if terms 1, 2 and 4 can be explicitly accounted for, then P_R is proportional to the backscatter cross-section:

$$\frac{\rho A}{\Omega R^2} \tag{3.5}$$

In the condition where $A = 1$ and the pulse is intercepted normal to a Lambertian target (i.e., $\Omega = 1$), an estimate of target ρ can be made through the radiometric calibration of range resolved P_R. In field conditions, it is rare for these assumptions to be met, so any estimate of ρ is often termed apparent reflectance, ρ_{app} (Jupp et al. 2009).

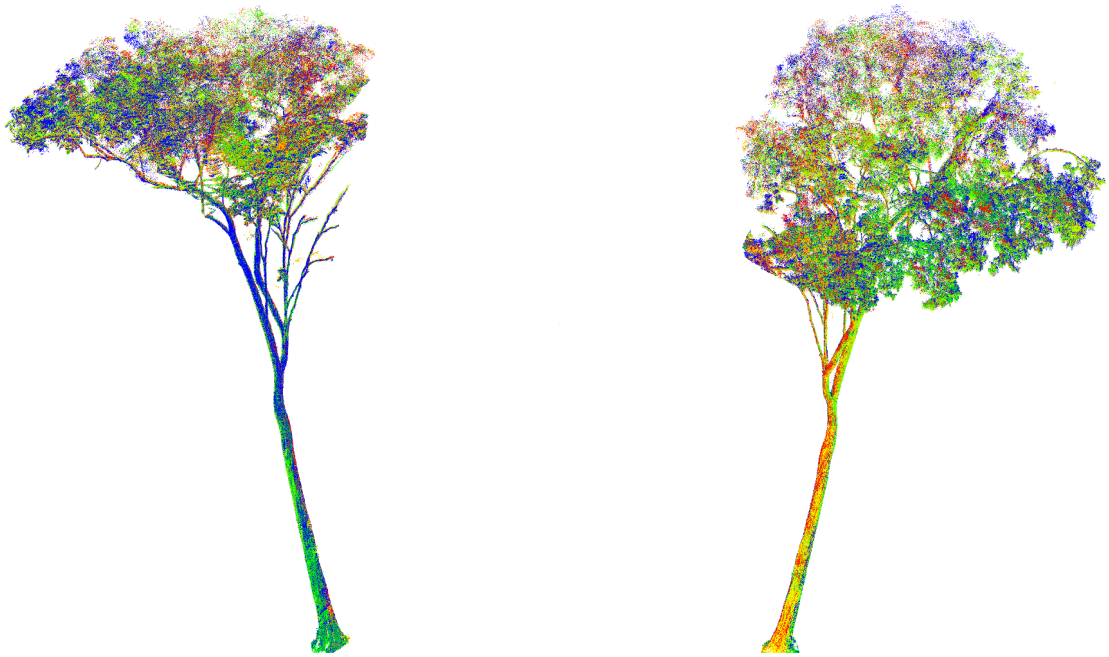
The radiometric calibration of a TLS instrument is non-trivial. This is in part due to variance in the characteristics of laser performance over time (i.e., temporal variance in η_S). A range of methods for radiometric calibration have been employed, from semi-empirical model fitting to neural networks and lookup tables (Li et al. 2016; Schofield et al. 2016; Pfennigbauer et al. 2010).

3.1.4 Scan co-registration

Considered up to this point, an individual scan, as shown in figure 3.1, provides a 3D sampling of the scene from the perspective of the scanner. Limited to this viewpoint, occluded objects in the scene cannot be sampled. To generate a full 3D sampling of the scene, data must be acquired from multiple viewpoints. These individual scans can then be mapped onto a common coordinate system to construct a scene-level point cloud.[1] This process is referred to as registration and is demonstrated in figure 3.4, where a tree-level point cloud representing a tropical forest canopy tree has been extracted from a plot-level point cloud and coloured by scan number ($1 - 121$).

For applications looking to exploit the rich information content of these 3D data,

[1] Referred to from now on as plot-level point cloud, as the interest in this thesis lies with scene-level point clouds of forest plots.



(a) (b)

Figure 3.4: A tree-level point cloud from a tropical canopy tree, extracted from the plot-level point cloud; coloured by scan number and viewed from two perspectives.

the accuracy of scan co-registration needs to be of the order of the accuracy of R. This is usually achieved by the strategic placement of markers through the scene prior to scanning. These markers will be uniquely identifiable from the rest of the scene in post-processing. Co-registration is then undertaken through correspondences estimation (i.e., locating common markers between scans and rotating/translating the point clouds until these markers occupy the same 3D space) (Fan et al. 2015). Alternatively, an active research area is scan co-registration without markers; instead, through feature extraction to create common tie points from objects already present in the scene (Kang et al. 2009; Weinmann et al. 2011), although this is usually at the expense of accuracy.

3.1.5 Data acquisition protocol

The quality of plot-level point clouds is defined here as the completeness with which the plot-level point cloud captures the 3D structure of the scene, within the bounds of fixed instrument-specific parameters (i.e., ranging accuracy and beam divergence).

Quality is determined by two user-defined parameters; firstly, the angular resolution, θ, determines the angular step between sequentially fired pulses in both azimuth and zenith. Figure 3.5 presents an idealised scenario with two infinitesimal, collimated pulses fired along the same zenith ring towards an extended target at range R.

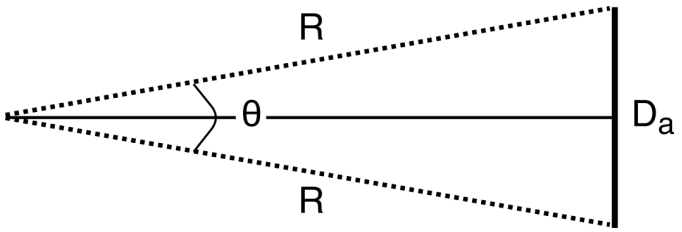


Figure 3.5: Two infinitesimal, collimated pulses (dashed lines) fired towards an extended target at range R, separated by the angular step, θ.

The distance between the location at where the two pulses intercept the target, D_a, in this idealised scenario is defined as:

$$D_a = 2R \tan\left(\frac{\theta}{2}\right) \tag{3.6}$$

This can be extended to account for beam spread at the target range as:

$$D_a = 2R \tan\left(\frac{\theta}{2}\right) - [D_E + R \tan \Theta] \tag{3.7}$$

In an idealised scene (e.g., the inside of a sphere with the scanner positioned at the centre, such that R is target invariant), D_a can be used directly in the definition of an appropriate value of θ. That is, if D_a is positive or negative, then the scene will be undersampled and oversampled respectively. However, in complex 3D scenes, this relationship is more uncertain as it would be expected for the distribution of target ranges to be highly variant. Selection of θ then, is usually determined on an ad-hoc basis through speculation as to the extent to which potential sources of occlusion in the scene will pollute these theoretical values.

The second parameter influencing data quality is the adopted sampling pattern (i.e., the location and number of scans used across the scene). These patterns can be either systematic or irregular, depending on the application. Systematic patterns include scan locations positioned through the scene in square, hexagonal or star formations with some fixed resolution. Irregular patterns are usually improvised to acquire sampling of a particular object in the scene (e.g., a single tree). Locations will be selected such that the other elements in the scene will interfere least in the sampling. The selection of sampling pattern is similar to θ. That is, user discretion and experience will determine the location and number of scans required to adequately sample the scene within the

time available.

3.1.6 The current state of TLS instruments

A broad and ever-increasing range of TLS instruments exist that target a variety of applications. Figure 3.6 presents a mix of five commercially-developed and research oriented instruments.

Figure 3.6: Left to right: RIEGL VZ-400, Dual Wavelength Echidna LiDAR, Canopy Biomass LiDAR, Salford Advanced Laser Canopy Analyser and FARO Focus 3D

The development of the current crop of high-end (.ca > £100k) TLS instruments has almost exclusively been driven by the requirement for acquisition of non-contact, surveying-grade data for applications across civil engineering, mining and policing. As a result, the hallmarks of these instruments are high accuracy and precision in the estimation of R (1–5 mm), low beam divergence (0.1–0.5 mrad) and high effective measurement rates ($>$ 100 kHz). These instruments will usually record the returning waveform and have the ability to resolve multiple targets per pulse. To exclude ambiguity in the source of returning radiation (i.e., due to scattering from a multiple target or multiple scattering from a single target) the instruments are usually characterised by a narrow detector field of view. As the accuracy in the estimate of R is paramount to these instruments, they are also often characterised by a long pulse length and small t_s, to aid in the accurate and precise decomposition of the waveform (Landi 2002). These scanners will often

operate in the near-infrared wavelengths, where the use of increased pulse energy to resolve low-reflectivity targets or longer ranges, will not compromise eye-safety. At these wavelengths, the reflectance of expected target materials, such as concrete, as shown in figure 3.2, remain relatively high.

In the mid-end range (.ca > £30k), there has been significant uptake in phase-shift instruments. An alternative to the time-of-flight method, phase-shift instruments constantly emit a pulse modulated at several frequencies. R is determined by the difference, the phase shift, in the frequency of the returning pulse. These instruments can be identified by very high effective measurements rates (> 1000 kHz) and high accuracy in their estimation of R (1–5 mm). However, the modulation process limits the maximum resolvable range, often to as low as 15 m.

There has been a move in recent years towards lower cost (.ca < £20k), smaller and more portable TLS instruments (i.e., not tripod mounted). These include the GeoSLAM ZEB-REVO, a hand-held scanner that continuously collects data whilst on the move. Of significant note is that co-registration is from the data themselves. That is, there is no requirement for the aforementioned co-registration targets. However, this technology is significantly less mature than the previously described instruments, so the accuracy in the estimation of R and the co-registration currently exclude them from precision 3D measurement. Additionally, a limited outdoor range of 15 m reduces the possible applications to forestry.

In the environmental sciences research domain, two operational TLS instruments have been developed explicitly for forestry applications. The Salford Advanced Laser Canopy Analyser (SALCA) and the Dual-Wavelength Echinda LiDAR (DWEL) are both instruments operating two lasers at wavelengths of approximately 1060 nm and 1540 nm (Danson et al. 2014; Douglas et al. 2015). They have both been developed on the premise that their radiometric calibration will provide estimation of target ρ_{app} at both wavelengths. As illustrated for the *Quercus spp.* bark and leaf in figure 3.2, variance in their reflectance at these wavelengths would allow their discrimination. This is shown in figure 3.7, where a point cloud acquired by the DWEL instrument in *Eucalyptus* spp. open forest in Tumbarumba, Australia, is coloured by the ratio of ρ_{app} at both wavelengths (Newnham et al. 2015). It can be seen that this ratio is capable of clearly distinguishing between wood and leaf targets. The potential application for this is reviewed in section

3.2.1.

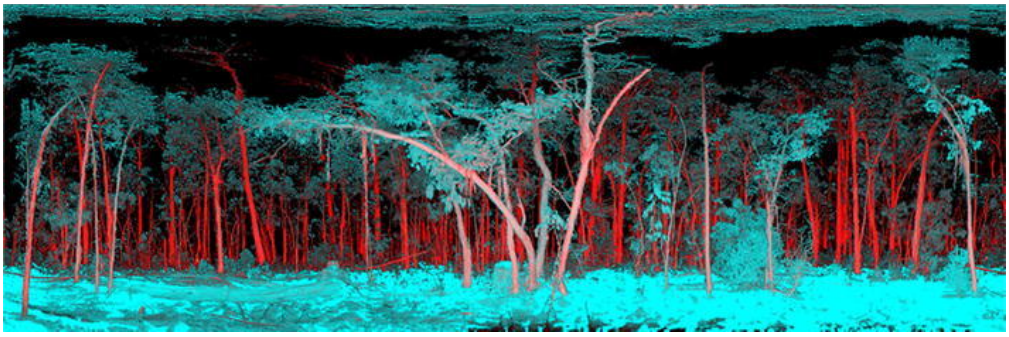


Figure 3.7: Point cloud acquired by the DWEL instrument from *Eucalyptus* spp. open forest in Tumbarumba, Australia; coloured by the ratio in ρ_{app} at wavelengths of 1060 nm and 1540 nm (Source: Newnham et al. 2015).

3.2 Measuring forest structure with TLS

This section reviews the body of published literature that describes retrieval of forest structure from TLS data. Section 3.2.1 briefly reviews work that has focused on plot-scale retrieval of forest structure. Section 3.2.2 then evaluates work, closely aligned to the objectives of this thesis, that aimed to retrieve tree-scale AGB from TLS data.

3.2.1 Plot-scale retrieval of forest structure

The majority of the studies that have focused on the plot-scale have derived parameters of forest structure from TLS data through an assessment of forest gap fraction. Gap fraction is described as the proportion of observable sky (i.e., not occluded by elements of the forest) as a function of view field and direction. An important parameter of forest structure, gap fraction is often used to abstract the complex 3D arrangement of forest elements into a single parameter (Welles et al. 1996). This can allow inferences to be drawn about the effect of the 3D arrangement of forest structure, in relation to a range of research questions. These include description of the canopy radiation regime, detection of anthropogenic disturbances and rainfall deposition rates (Widlowski et al. 2005; Peres et al. 2006; Pypker et al. 2005). Perhaps of most note, simplistic light models which use gap fraction can be inverted (Bréda 2003), to derive an estimate of the key biophysical property, leaf area index, LAI, defined as the one-sided leaf area per unit ground (Carlson et al. 1997). This is significant because of the correlation of both LAI, and the related radiometric property, fraction of absorbed photosynthetically active radiation, with primary production (Gower et al. 1999).

In the context of TLS data, gap fraction can be determined by the proportion of outgoing pulses that record a hit against those that do not. This is most usually expressed as gap probability, P_{gap}, such that when considering a single pulse from a first return instrument, then $P_{gap} = 0$ if the pulse is intercepted by a forest element, whilst $P_{gap} = 1$ if it is not. P_{gap} is typically quantified as a function of zenith angle, θ, to estimate $P_{gap}(\theta)$ for outgoing pulses across the azimuth range, and further, is often height-resolved (i.e., $P_{gap}(\theta, z)$) as (Lovell et al. 2003):

$$P_{gap}(\theta, z) = 1 - \frac{\#r | r\cos\theta < z}{N(\theta)} \tag{3.8}$$

Where, $\#r$, denotes the number of hits, proportional to the number of fired pulses, $N(\theta)$.

Jupp et al. (2009) derived a method for determining plant area index, PAI, (i.e., the plant area per unit ground), from this vertically resolved gap probability, by assuming a foliage orientation function, $G(\theta)$, as:

$$PAI(z) = \frac{-\ln(P_{gap}(\theta, z))\cos\theta}{G(\theta)} \tag{3.9}$$

Where $G(\theta)$ is either described through a parametric model, or in the special 'hinge-angle' case where $\theta = 57.5°$, disregarded (Jupp et al. 2007).

Differentiating $PAI(z)$ with respect to height then, will provide a description of the vertically distributed arrangement of plant material, known as the plant area volume density, $PAVD(z)$:

$$PAVD(z) = \frac{d(PAI(z))}{dz} \tag{3.10}$$

Cuni-Sanchez et al. (2016) demonstrated the power of these methods by applying them to TLS data acquired through a succession of tropical forest types, from young colonising forest to old-growth forest in Lopé National Park, Gabon. Figure 3.8a shows photographs of the four forest types, followed by figure 3.8b that presents estimates of $PAVD(z)$ for each of the forest types.

It can be observed that these plant profiles are readily capable of describing some of the vertical structure of the forest. For instance, the profile for the old-growth forest (F2) suggests the majority of the plot plant material is contained in the canopy layer,

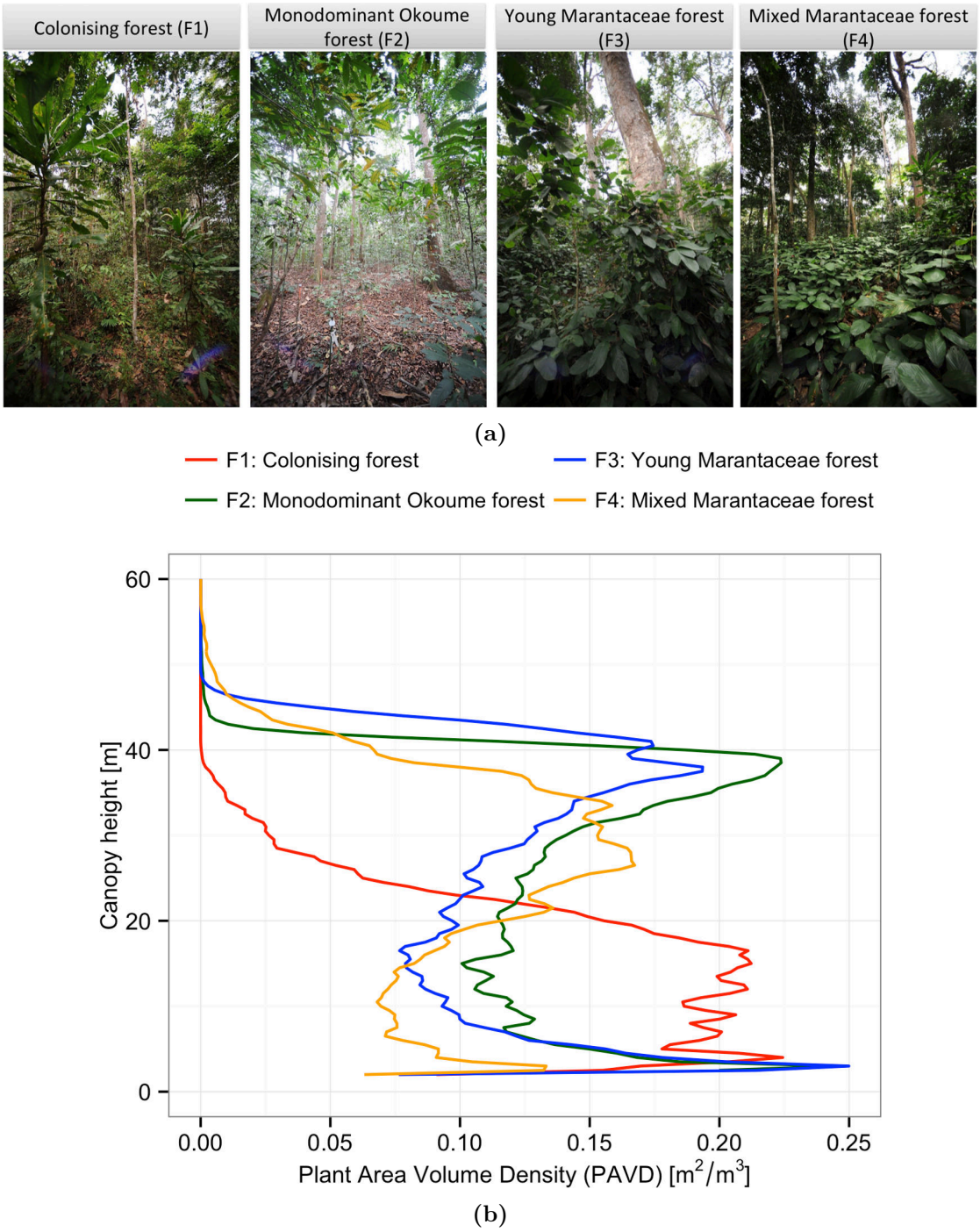


Figure 3.8: (b) Estimation of $PAVD(z)$ from TLS data acquired across a succession of the four forest types illustrated in (a) (Source for both images: Cuni-Sanchez et al. 2016).

in contrast to the young colonising forest (F1), where the majority of plant material is either competing in the understory to emerge, or is located in the undergrowth.

Recently, the methods described above for quantifying PAI and $PAVD(z)$ have been advanced to compensate for sensor configuration (including waveform instruments), topography, vegetation clumping and automation (Calders et al. 2014). As a result, these methods could be categorised as mature, with the retrieval of vertically resolved plant

profiles of forests plots being readily accessible to researchers looking to address research questions through this description of forest structure.

The next step for these methods is to transform the estimates of PAI into LAI. This is being advanced through the dual wavelength research instruments described in section 3.1.5. As illustrated in figure 3.7, the ability to discriminate between wood and leaf targets would allow the generation of leaf and wood only point clouds. These methods to estimate PAI can then be applied, as is, to the leaf only point cloud to produce an estimate of LAI. This will permit estimation of this crucial biophysical parameter to an accuracy that should far surpass that of traditional methods such as hemispherical photography.

3.2.2 Tree-scale estimates of above-ground biomass

This section reviews work aligned with the primary objective of this thesis: to demonstrate that TLS derived estimates of AGB can be both accurate and applied on a wide-scale (here meaning across tens of 1 ha forest plots).

Published methods for estimating tree-scale AGB from TLS data can be grouped into two distinct categories. The first set of methods, reviewed in section 3.2.2.1, describe approaches to retrieve traditional parameters of forest structure from TLS data, that are then used as allometric regressors. These parameters, such as bole diameter, D, and tree height, H, are ingested into allometric models, such as those described in chapter 2, to predict tree-scale AGB.

Alternatively, the second category of methods, reviewed in section 3.2.2.2, look to exploit the rich information content of TLS point clouds to near-completely characterise 3D tree structure through modelling approaches. These newer methods involve fitting geometric primitives to tree-level point clouds to reconstruct tree surface topography. Tree-scale AGB is then retrieved by estimating the volume of these geometric primitives, and then multiplying by average wood density. This section concludes by assessing the singular study that has employed these new methods to estimate tree-scale AGB across a forest plot.

3.2.2.1 Allometric regressors

One of the earliest published examples of using TLS data to derive estimates of the allometric regressors D and H was the study by Hopkinson et al. (2004). In the study,

TLS data were acquired from mixed species deciduous forest in Canada. During post-processing, tree-level point clouds were extracted from the plot-level point cloud through 'by hand extraction'[2]. Using these tree-level point clouds, estimates of D and H were obtained by measuring the Euclidean distance between key points.

A comparison between these TLS-derived structural parameters and their measurement via traditional methods (i.e., tape measure and inclinometer) identified that the TLS-derived estimates of D were both accurate and unbiased. TLS-derived estimates of H were found to be systematic underestimates. This was speculated to be a result of the adopted sampling regime.

This generic approach has been replicated by numerous studies across various forest types from Pine plantations in the UK, to mixed species broad-leaved deciduous forest in the USA (Watt et al. 2005; Tansey et al. 2009; Henning et al. 2006; Vastaranta et al. 2009). The results from these studies corroborated that D could be retrieved from TLS data both accurately and without bias. These studies concluded with the common view that TLS has the potential to revolutionise the measurement of forests.

The study of Maas et al. (2008) stands out as being one of the limited instances where tree-level point clouds were not obtained through 'by hand extraction'. Instead, point clouds representing individual stems were delineated through automatic point cloud manipulation techniques, such as spatial clustering[3]. D was then successfully determined to a high accuracy through automated circle fitting for 97 % of the assessed individuals.

In these aforementioned studies, any uncertainties in the estimate of D were attributed to either the adopted sampling regime or the instrument-specific accuracy in the estimation of R. The impact of these parameters on the accurate retrieval of D was demonstrated by the instrument inter-comparison study of Bauwens et al. (2016).

A limitation to all these studies is that the accuracy in their estimate of D and H did not surpass the accuracy of estimates made through traditional measurement. With the high costs associated with TLS, relative to the traditional methods, in both the instrument purchase and personnel training, it is difficult to justify the use of TLS when there is no material increase in the quality of the retrieved parameters. This justification

[2] 'By hand extraction' is used here and from now on to describe the manual manipulation of a point cloud. That is, the tree-level point cloud representing an individual tree is extracted from the plot-level point cloud through the sole discretion of the user as to which points in the plot-level point cloud exclusively belong to the individual tree in question.

[3] There are a number of generic point cloud manipulation techniques, including spatial clustering, that permit access to the information content of a point cloud; a number of them are described in chapter 5.

is harder still when the laborious and time-consuming effort required for the 'by hand extraction' is considered.

Perhaps a more fundamental limitation to tree-scale AGB estimation via TLS-derived allometric regressors is that the estimates are exposed to the same uncertainties as the estimates obtained via traditional measurement. That is, tree-scale allometric modelling uncertainty, as defined in chapter 2, is the same across both estimates. The only avenue available to these methods for reducing allometric uncertainty is to reduce the tree-scale measurement uncertainty component. Even if TLS-derived estimates could improve on the accuracy in the estimation of D and H beyond that permitted by the traditional measurement, it was demonstrated in chapter 2 that the measurement uncertainty component largely averages out when up-scaling to the plot; negating any possible benefit introduced by TLS data.

3.2.2.2 Quantitative structural models

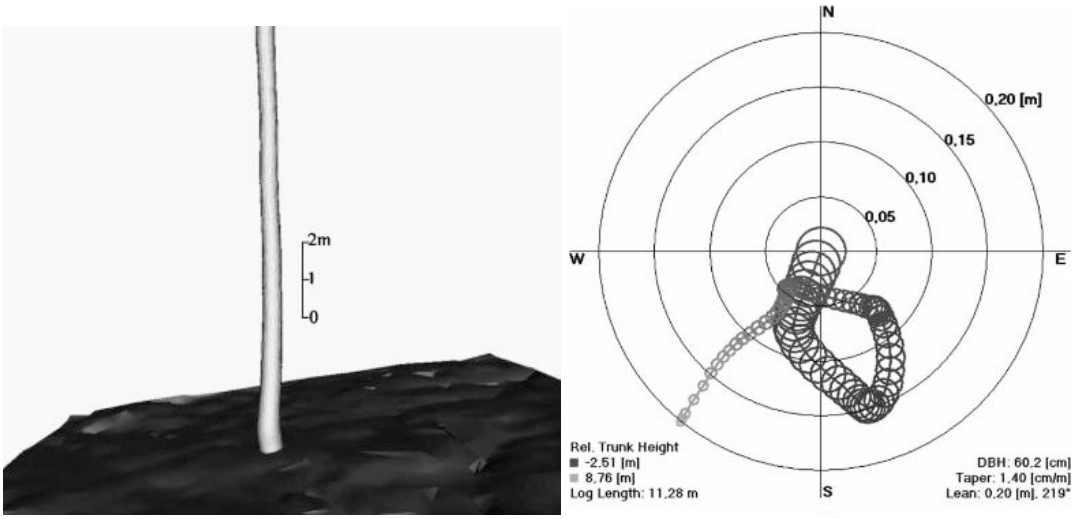
The experimental study presented in chapter 2 established that the allometric regessors of D and H were incapable of describing pan-tropical variance in AGB with any reasonable certainty. If TLS-derived methods are to provide more accurate and unbiased estimates of tree-scale AGB than that permitted by allometry, then it is necessary to obtain a more complex description of tree-scale structure from the data. The five studies reviewed in this section have achieved this by reconstructing the surface topography of part (e.g., the stem and or branching), or, all of the tree (i.e., including leafy material) from tree-level point clouds.

For the woody portion, this has universally involved fitting cylinders to subsections of the tree-level point cloud to construct a cylinder model that represents 3D tree structure, referred to from this point on as a 'quantitative structural model' (QSM). Cylinders are the geometric primitive of choice as they are most capable of describing woody tree architecture across the broadest range of scenarios (Åkerblom et al. 2015). Fitting a cylinder to a point cloud is most commonly undertaken through a random sample consensus (RANSAC) cylinder model fit[4].

An early study reconstructing tree surface topography from TLS data is the publication by Thies et al. (2004). In the study, the authors extracted two tree-level point

[4]This is another of the generic point cloud manipulation techniques that are described in chapter 5.

clouds using 'by hand extraction' from two plot-level point clouds acquired in European Beech and Wild Cherry forest plots. Multiple cylinders were then fitted through each stem to generate the stem model shown in figure 3.9a. From the model, an estimate of stem taper and lean, was derived from the radius and orientation of each fitted cylinder, as shown in figure 3.9b.



(a) **(b)**

Figure 3.9: (a) Stem model derived from fitting multiple cylinders to a tree-level point cloud of a European Beech tree, and, (b) a bird's-eye view of the position and radius of each cylinder in the model from which a description of taper and lean was obtained (Source for both images: Thies et al. 2004).

Dassot et al. (2012) expanded on the application of these QSMs by deriving estimates of stem volume from 42 individuals in Oak and Ash forest plots. Tree-level point clouds were obtained through 'by hand extraction' of the plot-level point cloud. For each tree-level point cloud, a polyline was then manually drawn through each subsection of the point cloud that represented a segment of stem or branching. Cylinders were manually expanded around each polyline until they encompassed their subsection of points. A tree-level point cloud and the subsequent QSM derived via these methods can be seen in figure 3.10.

Post-scanning, the 42 individuals were destructively harvested to provide method validation. However, direct measurement of stem volume was not made with these data, instead, stem volume was inferred from measurements of D every 2 m along the stem. It was shown that TLS-derived stem volume was, on average, estimated to within 10 % of the harvest-derived estimates. This small divergence was thought to arise from a systematic overestimation in the harvest-derived estimates.

Figure 3.10: Tree-level point cloud (left) and derived QSM (right), the manually inserted red polylines visible in the high order branching defined the position and orientation of each of the grown cylinder (Source: Dassot et al. 2012).

The study of Côté et al. (2009) remains unique as the objectives were to reconstruct tree surface topography in its entirety (i.e., the wood and leaf components). This was achieved through a novel amalgamation of TLS data and parametric modelling; a tree-level point cloud and subsequently derived QSM from these methods is shown in figure 3.11.

Once the tree-level point cloud had been derived via 'by hand extraction', the leaf and wood targets were discriminated through classification of ρ_{app}. A skeletal curve (Verroust et al. 2000) was then generated around prominent sections of woody structure from which cylinders were fitted. A set of consistency tests were then conducted to assert connectivity between the fitted cylinders. For cylinders representing the highest order branching, additional branching or leafy material, guided by the discriminated leaf targets, was appended to the QSM according to the parametric model.

The application of the QSMs in Côté et al. (2009) were to drive radiative transfer simulations across a variety of sensor configurations and wavelengths to aid development of new sensors and product retrieval algorithms. Hence, the QSM was required to faithfully represent the structure of the tree from which it was derived. Because of these objectives, the models were not developed for deriving an estimate of tree-scale volume, so their application to tree-scale AGB estimation was not considered. In addition, wide-spread application of this algorithm was limited by design. The parametric model determining the appendage of high order branching and leafy material made the algorithm species-specific. Discrimination by ρ_{app} also made the algorithm instrument-

specific.

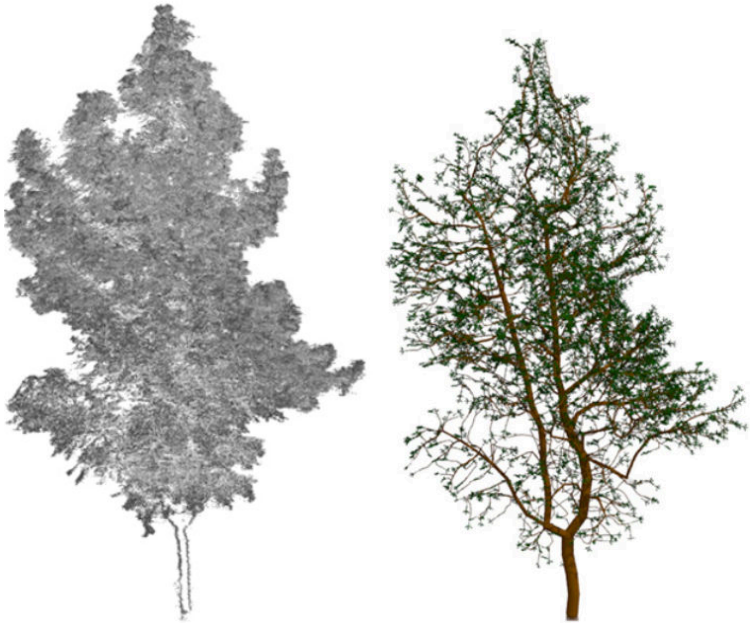


Figure 3.11: Tree-level point cloud (left) and derived QSM (right) constructed from a skeletal curve and cylinder fit, the high order branching and leafy material was appended to the QSM by a parametric model (Source: Côté et al. 2009).

The study by Raumonen et al. (2013) produced the first published methods for the full and automated reconstruction of woody structure from tree-level point clouds. Here, the tree-level point clouds were segmented into small patches with metrics such as the angle and distance between patches used as heuristics to delineate branching. Cylinders were then fitted through these delineated sections of branching[5]. Figure 3.12 presents a tree-level point cloud and subsequently derived QSM from these methods.

Method validation has been undertaken by using TLS data acquired through radiative transfer simulations driven by procedurally generated tree models, such that tree structure was known *a priori* (Disney et al. 2012). From these data, it was shown that the QSMs overestimated tree-scale volume by 2 %. Hackenberg et al. (2015) presented a similar automatic approach for the construction of QSMs representing full woody structure. Again using artificial data, it was shown that these QSMs overestimated tree-scale volume by 3 %.

Calders et al. (2015) produced the first, and to date, only, 'in the field' application of these methods across a forest plot. In this study, 65 individuals from *Eucalyptus* spp. open forest in Australia were scanned with tree-level point clouds obtained via 'by

[5]This method is extensively used in this thesis to derive estimates of tree-scale AGB. A full description of the method can be found in chapter 6.

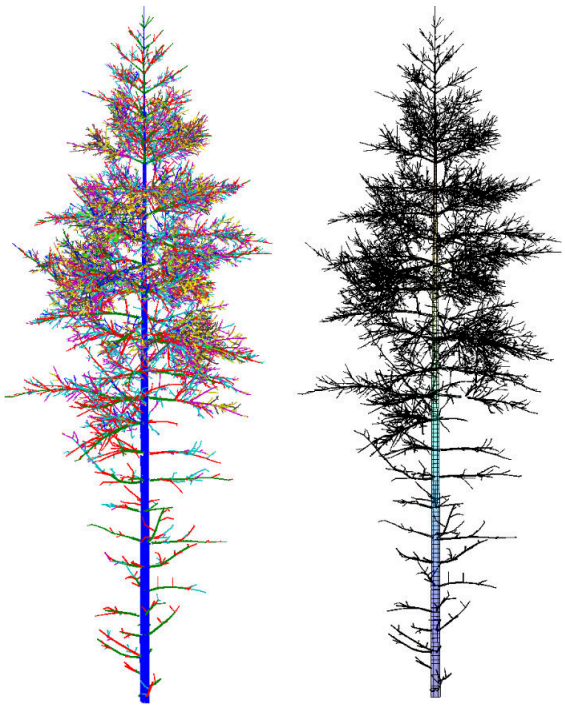


Figure 3.12: Tree-level point cloud (left) coloured by delineated branching and derived QSM (right) constructed by segmenting the point-cloud into small patches whose heuristics delineate branching to which cylinders are fitted (Source Raumonen et al. 2013).

hand extraction'. Post-scanning, the individuals were destructively harvested and direct measurement of tree-scale AGB was made. A comparison of the TLS-derived estimates of tree-scale AGB (here derived using the Raumonen et al. (2013) QSM algorithm alongside a constant value of wood density obtained from the destructive harvest) against the directly measured tree-scale AGB can be seen in figure 3.13. (Calders et al. 2015)

This comparison identified that the TLS-derived estimates of tree-scale AGB were, on average, 10 % larger than the destructive measurements. This was in contrast to the species-specific allometric model that, on average, underestimated tree-scale AGB by 30 %. Of particular note, Calders et al. (2015) showed that the overestimate in TLS-derived estimates of tree-scale AGB were independent of tree size. The results of this study, with the unique validation dataset, provided proof of concept that accurate estimates of tree-scale AGB could be derived from TLS data.

The five studies reviewed here proposing methods to construct QSMs, in particular, those of Raumonen et al. (2013) and Hackenberg et al. (2015) have demonstrated that TLS data can be used to construct complex descriptions of the topographic surface of woody tree structure. In these studies it was shown that QSMs derived from high quality TLS data permit the accurate retrieval of tree-scale volume. From this, as shown by the

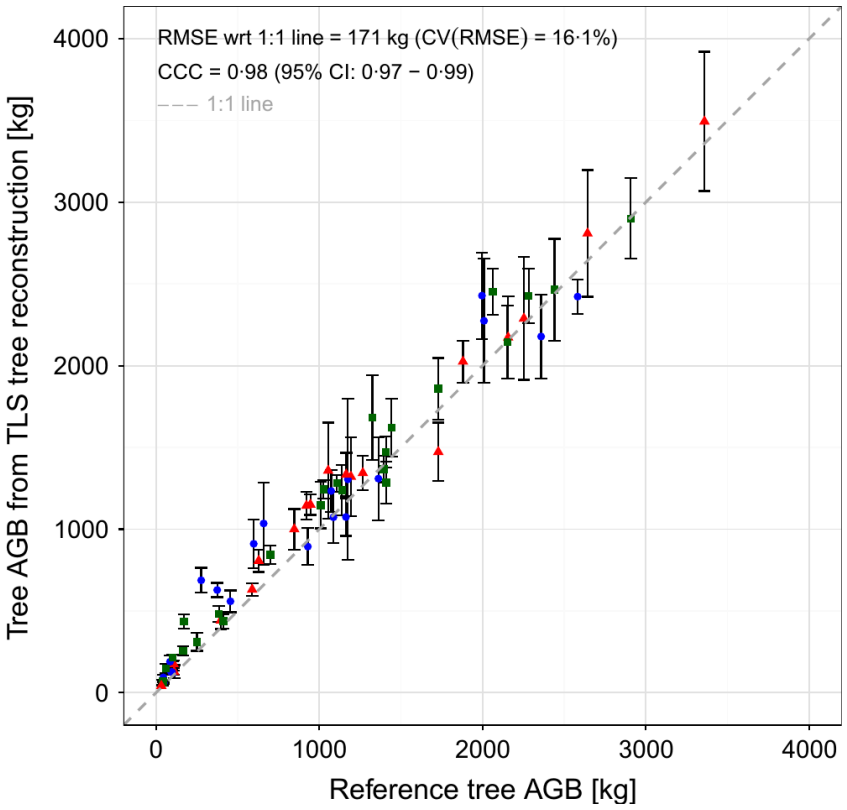


Figure 3.13: Comparison of tree-scale AGB for 65 *Eucalyptus* spp. individuals obtained from TLS-derived volume estimation and direct measurement (Source: Calders et al. 2015).

study of Calders et al. (2015), tree-scale AGB could be accurately estimated when accompanied by an estimate of wood density.

These TLS-derived estimates of tree-scale AGB via volume estimation are in stark contrast to the TLS-derived estimates of tree-scale AGB obtained via allometric modelling. The fundamental distinction is that there is no longer any reliance on inference from the allometric model and underpinning dataset to capture the influence of tree structure variability on AGB. Instead, the volume-derived estimate of AGB has used the information content of the point cloud to capture this variability at source. That is, this estimate is entirely independent of allometric modelling and the associated uncertainties that were demonstrated in chapter 2. As such this represents a potential step change in the way tree-scale AGB is measured.

The cause of the overestimates identified by the studies of Disney et al. (2012) and Calders et al. (2015), alongside any generic uncertainty attributed to these TLS-derived tree-scale estimates of volume and AGB can be assigned to five sources. First, with the exception of the instrument-specific algorithm of Côté et al. (2009), all returns in the tree-level point cloud are treated as equal. This presents significant potential for

leaf returns to interfere in the construction of the QSMs. This would most likely invoke an overestimation, as the additional points will potentially result in the generation of more, or larger cylinders. Second, the quality of the plot-level point cloud, influenced by the sampling regime and any data contamination from wind or precipitation, will likely impact on the goodness of the QSMs.

Third, the use of cylinders as the geometric primitive will most likely lead to a systematic overestimate of tree-scale volume. This is due to localised branch tapering that would perhaps be better described by a bi-truncated conic section. Fourth, alongside the tree-level point cloud, each of the five models have additional inputs (e.g., in the Raumonen et al. (2013) algorithm, the definition of the size of the patches) that exert a large influence on the goodness of the resultant QSMs. Such model inputs are often abstracted, and as such are unintuitive and potentially subjective, making their definition and subsequent assessment on the accuracy of the retrieved estimate of volume, non-trivial. Finally, the use of a constant density value across the tree will introduce some uncertainty in estimates of tree-scale AGB as wood density is radially, height and species variant (Knapic et al. 2008; Nock et al. 2009).

3.3 Conclusions

Chapter 2 presented a novel study that demonstrated a requirement for the introduction of alternative methods to allometry for the estimation of tree- and plot-scale AGB. In this chapter a review of the alternative method considered in this thesis, TLS, has been undertaken. This began with an introduction to the principles of TLS and LiDAR. It was shown that commercially available TLS instruments are capable of acquiring a rapid, accurate and near-complete sampling of the 3D structure of a forest scene.

A review of the published literature whose objectives aimed to retrieve parameters of forest structure from TLS data was then presented. It was shown, that at the plot-level, descriptions of the complex arrangement of vertical forest structure can be derived from methods estimating gap fraction. An assessment of the published work aiming to retrieve tree-scale AGB from TLS data was then undertaken. It was shown that there are two distinct sets of methods presented in the literature for retrieving tree-scale AGB from tree-level point clouds; these are:

1) To retrieve estimates of the allometric regressors (e.g., D and H) from the tree-level

point cloud, and estimate tree-scale AGB via allometric modelling.

2) To fit geometric primitives to the tree-level point cloud to reconstruct tree surface topography, and estimate tree-scale AGB via volume estimation.

Chapter 2 illustrated that the allometric regressors D and H are incapable of describing pan-tropical variance in AGB. Therefore, the only improvement that can be offered by estimates of tree-scale AGB derived via method 1), is to provide a more accurate estimation of the allometric regressors than that permitted by traditional measurements (e.g., tape measure). Even if this was achieved, it was demonstrated in chapter 2 that when up-scaling tree-scale estimates of AGB to the plot-scale, the uncertainty associated with the measurement of allometric regressors largely averaged out. Method 1) then, cannot be expected to offer any improvements over traditional methods.

Five studies falling into the second category were then reviewed, of particular note were Raumonen et al. (2013) and Hackenberg et al. (2015), whose algorithms were shown to permit accurate reconstruction of woody structure. This explicit definition of tree structure was derived exclusively from the tree-level point cloud. As such there is no requirement for inference of the influence of tree structure variability on AGB, as this variability has been caught at source. These methods then, are entirely independent of allometry. The single dependency of these methods in the estimation of tree-scale AGB is for an accompanying estimate of wood volume.

Using artificial data with an *a priori* definition of tree structure, these methods were shown to reproduce tree-scale volume to within a few percent. The study of Calders et al. (2015) provided further proof of concept by demonstrating the use of these methods on real-world data. With destructive harvest data acting as validation, it was shown that estimates of TLS-derived tree-scale AGB via volume estimation could be made to within 10 % of the direct measurement. This is in contrast to total tree-scale allometric uncertainty that was shown in chapter 2 to exceed 75 %.

These studies have demonstrated the potential paradigm shift offered by estimates of tree-scale AGB derived via method 2), as they have been shown to be significantly more accurate than, and entirely independent of, allometry. However, application of these new methods needs to be practically implementable on wide-scale (here meaning across tens of 1 ha forest plots) to make them a viable alternative to allometry. The review of these

new methods in this chapter has highlighted the following knowledge gaps that currently prohibit wide-scale uptake:

i) The automatic extraction of individual tree-level point clouds from plot-level point clouds.

ii) The automatic inference of tree-scale AGB from these individual tree-level point clouds.

iii) The quantification of uncertainties that are introduced by these new methods.

Knowledge gap i), the extraction of tree-level point clouds from plot-level point clouds has been a recurring theme throughout section 3.2.2. Near-universally, 'by hand extraction' has been used throughout the literature to extract tree-level point clouds. This approach is time consuming, and across multiple forest plots is simply impractical. This impracticality is compounded when applied to TLS data acquired in more structurally complex forest types, such as tropical forest. Development of automatic methods to achieve extraction of tree-level point clouds is an essential requirement to the wide-scale implementation of forest structure retrieval from TLS data. Chapter 5 presents *treeseg*, an algorithm that has been developed during the work for this thesis to bridge this knowledge gap.

Knowledge gap ii) also concerns automation. It was highlighted in the review that the goodness of the QSMs, from which tree-scale AGB estimates are inferred, are dependant on a series of input parameters other than the tree-level point cloud. In the instance of the Raumonen et al. (2013) algorithm, these input parameters play a significant role in the accurate reconstruction of structure and subsequent volume estimation. As these parameters vary significantly on a tree-level point cloud by tree-level point cloud basis, the automatic selection of these parameters is essential to the wide-scale implementation of these new methods. Chapter 6 presents a theoretical framework developed during the work for this thesis to bridge knowledge gap ii).

Knowledge gap iii) is the quantification of uncertainties associated with these new methods. Chapter 2 provided a novel framework to quantify the sources of uncertainty, and the magnitude of their impact on allometric modelling. A rigorous understanding in the uncertainties of these TLS-derived methods is required if they are to become viable alternatives to allometry. Knowledge gap iii) is bridged in chapter 6 by explicitly

quantifying the error in TLS-derived estimates of AGB through validation data acquired across a number of trees that have been destructively harvested post-scanning to obtain direct measurement of AGB.

Chapter 7 concludes this thesis by demonstrating these knowledge gaps have been bridged by applying these new methods to the TLS data described in chapter 4, that have been acquired across several forest plots. From these data, 441 individuals from 3.25 ha of predominately tropical forest are automatically extracted and reconstructed to retrieve estimates of tree- and plot-scale AGB.

4 Description of the field measurements

Chapter 2 presented an experimental study demonstrating that uncertainty in pan-tropical estimates of above-ground biomass (AGB) exceeds 75 % and 25 % at the tree- and plot-scale respectively. Chapter 3 provided a review of terrestrial laser scanning (TLS), the alternative method considered in this thesis to allometry for deriving more accurate estimates of tree- and plot-scale AGB. Chapter 3 concluded by identifying three knowledge gaps that currently prohibit tree- and plot-scale AGB retrieval from TLS data on a wide-scale (here meaning deployment across tens of 1 ha forest plots). Chapters 5 and 6 present experimental studies developing and applying new algorithms to bridge these knowledge gaps and to demonstrate that tree- and plot-scale AGB can be accurately retrieved from TLS data. This chapter describes the TLS data that has been collected during the work for this thesis, from which these subsequent studies have been conducted.

This chapter begins with an introduction in section 4.1 to the RIEGL VZ-400 terrestrial laser scanner, a time-of-flight, waveform instrument that has been deployed to capture all the data described in this chapter. Subsection 4.1.1 reviews the instrument parameters, followed by subsection 4.1.2 describing the decomposition of the waveform to derive estimates of range. Next, subsection 4.1.3 outlines the radiometric calibration of the instrument. Subsection 4.1.5 presents the adopted data acquisition protocols. Post-processing methods for scan co-registration and noise reduction are then reviewed in subsection 4.1.4 and 4.1.6 respectively. Finally, subsection 4.1.7 describes potential contamination of the data.

Section 4.2 then details the field sites at which TLS data have been captured. This is broken into: Karawatha Forest Park, Australia (4.2.1), Lopé National Park, Gabon (4.2.2), Alice Holt Forest, UK (4.2.3), Caxiuanã National Forest, Brazil (4.2.4) and Nouragues Nature Reserve, French Guiana (4.2.5).

4.1 The REIGL VZ-400 terrestrial laser scanner

The RIEGL VZ-400 terrestrial laser scanner is a commercially available instrument, generically targeted towards applications requiring acquisition of non-contact, surveying-

grade, 3D information of an object or scene. Broadly, the instrument design remit is to permit continuous collection of very high quality and rapid, estimates of range, in adverse environmental conditions, whilst maintaining a reasonable degree of portability.

Figure 4.1 shows the instrument deployed in the tropical forests of Lopé National Park, Gabon. Figure 4.2 then presents a point cloud acquired by the instrument in the tropical forests of Caixuanã National Forest, Brazil. This point cloud, coloured by 'relative reflectance' (described in subsection 4.1.3), ρ_{rel}, consists of 27 755 657 range estimates digitised to within an accuracy of 5 mm from 22 358 510 emitted pulses during a single 3 minute scan.

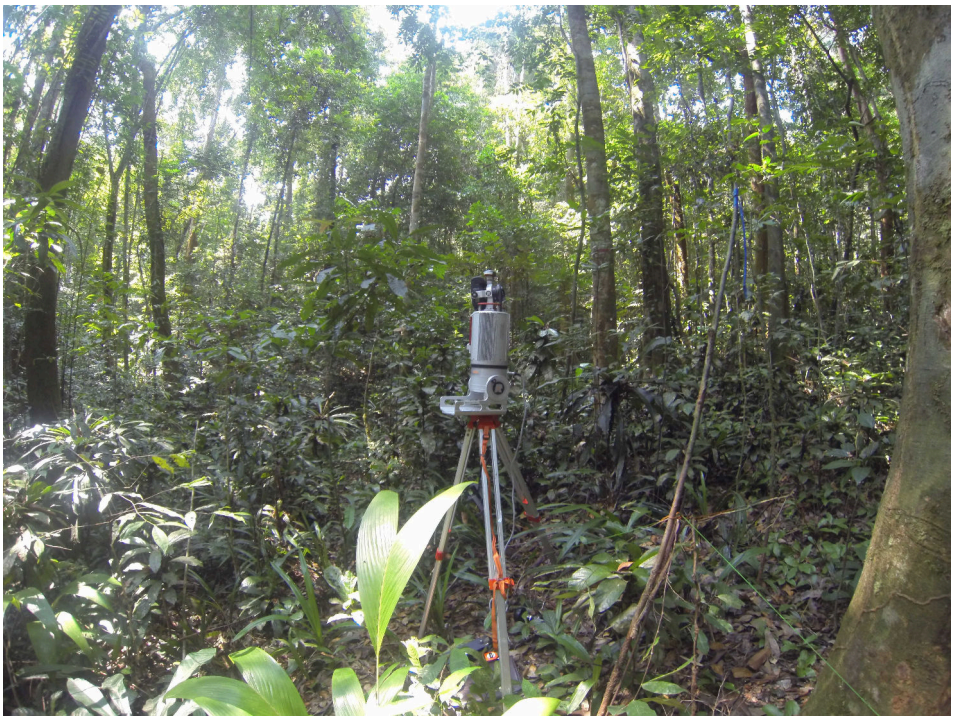


Figure 4.1: The RIEGL VZ-400 terrestrial laser scanner deployed in tropical forest in Lopé National Park, Gabon.

Figure 4.2: Panoramic view, from the perspective of the instrument, of a point cloud acquired by the RIEGL VZ-400 in tropical forest in Caixuanã National Forest, Brazil; coloured by ρ_{rel}.

4.1.1 Instrument parameters

Table 4.1 defines the characteristics of the RIEGL VZ-400 terrestrial laser scanner.

Ranging method	Time-of-flight
Wavelength	1550 nm
Angular resolution	$\geq 0.0024°$
Azimuth field of view	0° to 360°
Zenith field of view	−40° to 60°
Beam divergence	0.35 mrad
Beam diameter at emission	7 mm
Beam diameter at 100 m	42 mm
Pulse energy	0.48 µJ
Pulse length	3 ns
Sampling interval	1 ns
Peak pulse repetition rate	300 kHz
Effective measurement rate	122 000 meas s^{-1}
Ranging accuracy	5 mm
Ranging precision	3 mm
Minimum range	1.5 m
Maximum range	160 m @ $\rho = 20\%$
Resolvable targets per pulse	4
Minimum distance between multiple targets	0.8 m
Weight	9.6 kg
Battery chemistry	NiMH
Battery lifetime	4 h
Weatherproofing	IP64
Laser safety classification	1 (Eye safe)

Table 4.1: Parameters of the RIEGL VZ-400 terrestrial laser scanner.

As described in section 3.1.1, the instrument ranging accuracy and the diameter of the beam at the target, define the smallest resolvable distance. Notably, at a range of 60 m (a typical range between the instrument and top of a 40 m forest canopy), the beam diameter is 28 mm. The combination of these two characteristics make the RIEGL VZ-400 well suited to the 3D sampling of tropical forests with a resolution of the order of high order canopy branching. The ability to resolve up to four targets per pulse from the waveform also aids in the sampling of these complex scenes by reducing the influence of occlusion on data acquisition. Finally, the robust construction, low power requirement and light weight, make the instrument particularly well suited to the challenging conditions of tropical forest.

4.1.2 Ranging and waveform decomposition

As described in section 3.1.2, the estimation of range from instrument to target, R, in waveform instruments, is derived from the decomposition of the returning waveform. Figure 4.3a presents a small segmented section of the point cloud from figure 4.2. Three targets have been identified in this point cloud: 1) a return from the centre of a tree stem oriented normal to the scanner, 2) a return from a leaf oriented normal to the scanner, and 3) a return from a leaf oriented horizontal to the scanner. The waveforms from which these three targets have been digitised are shown in figure 4.3b.

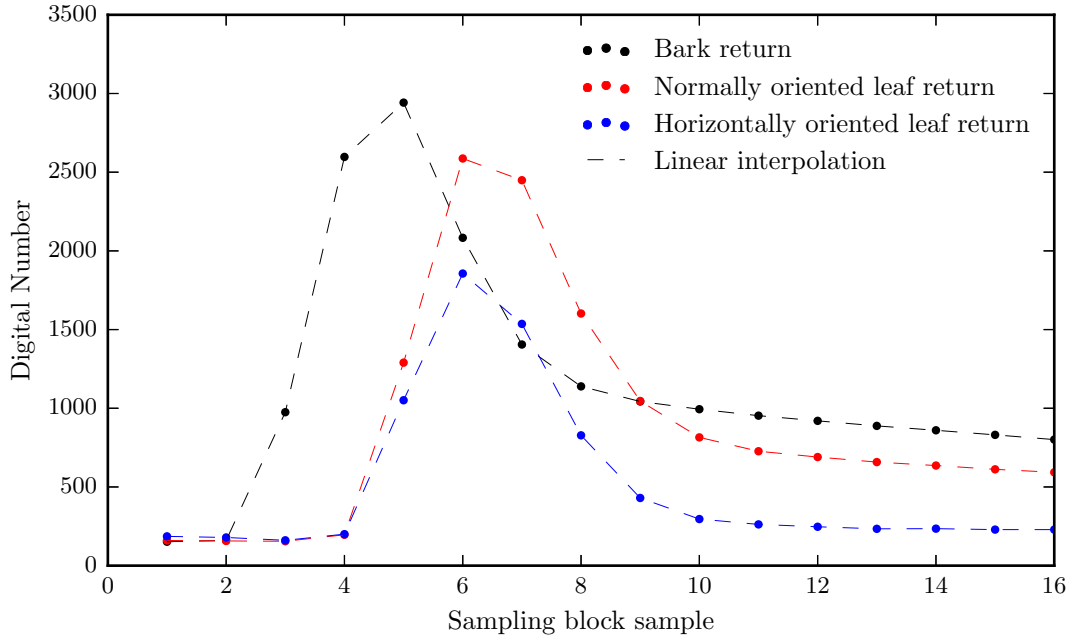
The most notable aspect of figure 4.3b is the infrequency of the waveform sampling. A sampling interval of 1 ns, corresponds to a sample of the returning pulse every 15.00 cm. To achieve a ranging accuracy of 5 mm with this relatively course sampling interval, a long pulse length is required, here three times greater than the sampling interval. This combination of sampling interval and pulse length contribute to the high effective measurement rate of the instrument ($122\,000\,\mathrm{meas\,s^{-1}}$), but, at the sacrifice of a minimum resolvable distance between multiple targets of 0.8 m. Of additional note is that none of the three signals in figure 4.3b return to the noise floor. The reason for this is not known but is speculated to be a result of detector saturation.

Uniquely, the instrument decomposes the waveform on the fly using on-board firmware (i.e., the point cloud is constructed in 'real time' and not as a post-processing activity). The methods used in the decomposition of the waveform are proprietary, with the particulars unknown, although a description at the high-level is provided by Pfennigbauer et al. (2010).

The strategy centres around the construction, in the manufactures test facility, of an instrument-specific calibration database of waveforms that fully characterise the dynamic behaviour of the instrument (i.e., in both the signal amplitude and temporal domains). In a field situation then, an incoming waveform will be matched to the closest representative in the database, where some further decomposition method will be employed to estimate R. A significant benefit over more routinely encountered methods such as a catch-all non-linear least squares fit is that instrument and situation-specific system attenuation can be largely eliminated; contributing to the high ranging accuracy of the instrument.

Whilst the full waveform is sampled by the instrument, the majority of the data is

(a)



(b)

Figure 4.3: (a) Section of point cloud segmented from the point cloud in figure 4.2, highlighting three targets: 1) a return from the centre of a tree stem oriented normal to the scanner, 2) a return from a leaf oriented normal to the scanner and 3) a return from a leaf oriented horizontal to the scanner; (b) the waveform from which each of these three targets have been resolved.

discarded post-decomposition. Only the sampling blocks from which a resolved target is digitised can be optionally stored; this is due to the expectation that the interest of the end user lies solely with the derived point cloud.

Per resolved target, the manufacturer provides a metric, referred to as 'deviation', that acts as a proxy to the goodness of the estimate of R. This value describes the

resemblance between the form of the returning waveform to the form of the matched waveform from the calibration database. Figure 4.4 presents five waveforms from which a target has been digitised, whose deviation values range from 0 to 100 (the dynamic range is 256). It can be seen that as the deviation value tends towards larger values, there is an associated increase in the full width half maximum of the signal.

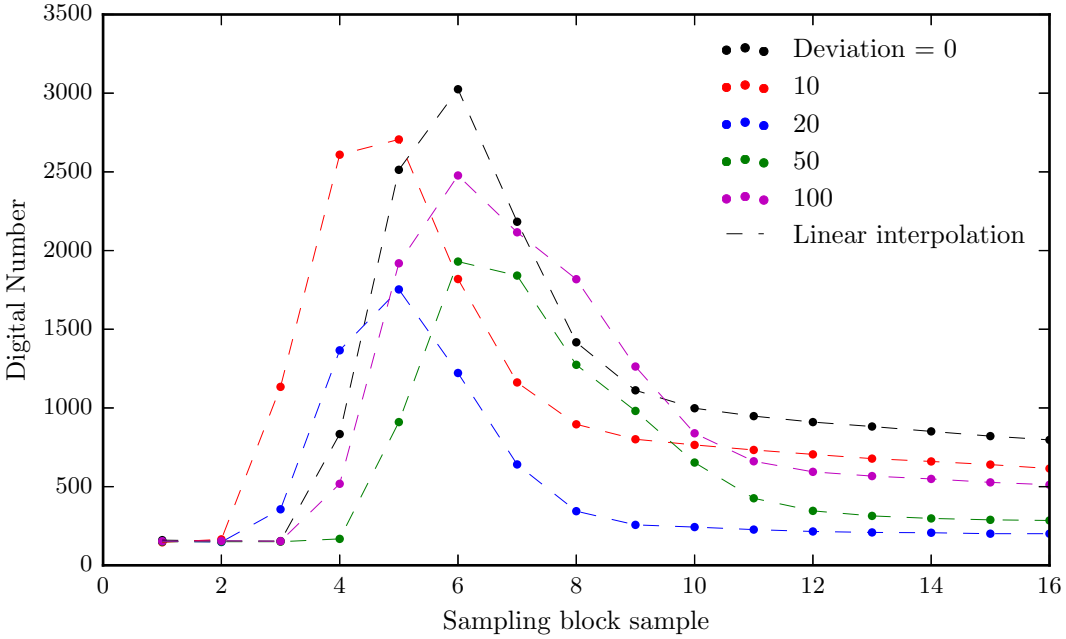
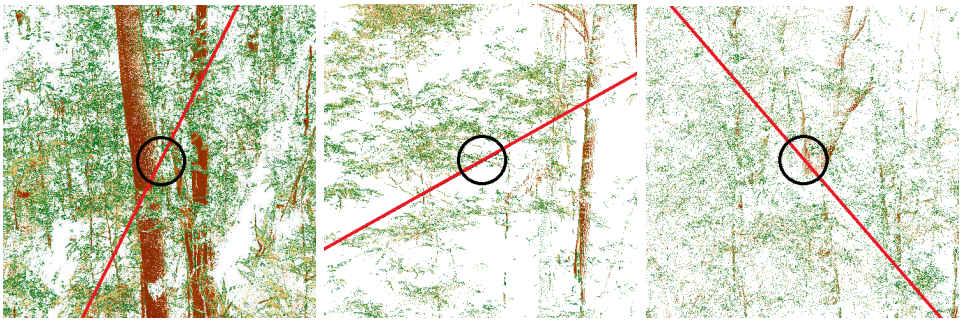


Figure 4.4: Five waveforms from which five targets have been digitised, whose deviation metric, a proxy to the goodness of the estimate of R, ranges from 0 to 100

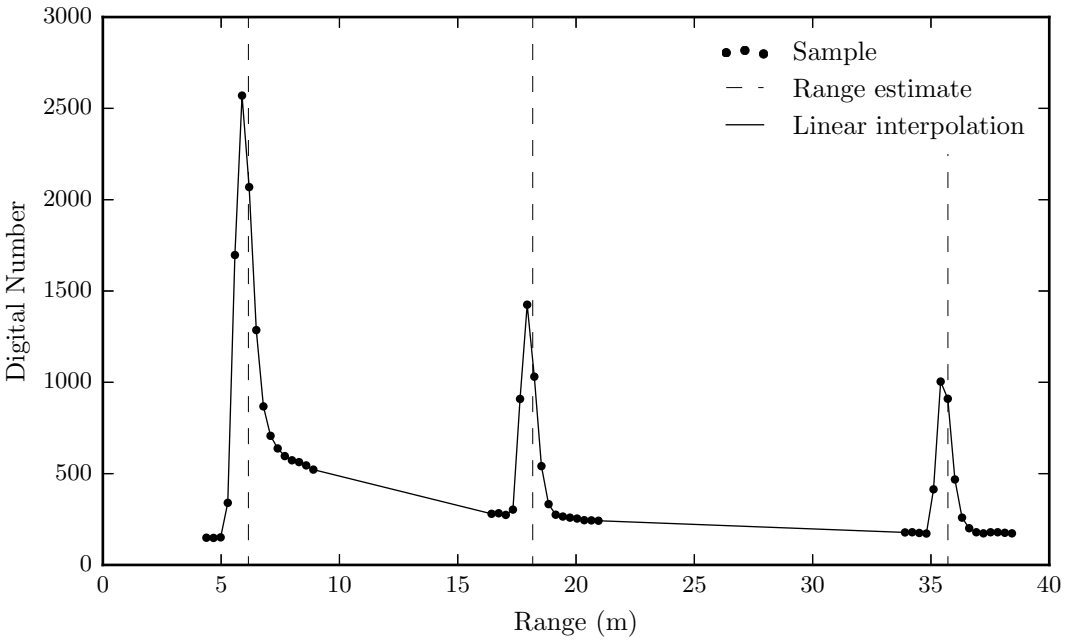
Per pulse, a manufacturer-imposed maximum of four targets can be digitised from the waveform. Figure 4.5a illustrates the travel of a single pulse through a section of the point cloud from figure 4.2, from which three targets have been digitised. The footprint of this selected pulse partially intersects: the edge of a stem in figure 4.5b, followed by a clump of leafy material in figure 4.5c, and finally some branching in figure 4.5d. Figure 4.5e then presents the returning waveform from this pulse.

The distribution of returns by return number (i.e., 1 is a first return; 2 , 3 and 4 represent multiple returns) across the point cloud shown in figure 4.2 is presented in figure 4.6. In total, multiple returns contribute in excess of 20 % of all returns; this is broadly consistent with all data acquired during the work for this thesis in tropical forest. This is also typical at the tree-level, as demonstrated in figure 4.7, where a tree-level point cloud representing a dominant canopy tree from tropical forest in Nouragues Nature Reserve, French Guiana has been segmented from the plot-level point cloud. In

(a)



(b) (c) (d)



(e)

Figure 4.5: Travel of a single pulse (coloured red) through (a) a segmented section of the point cloud from figure 4.2, representing a tropical forest canopy, with the pulse partially intersecting (b) the edge of a stem, (c) a clump of leafy material and (d) branching; where (e) shows the returning waveform from which the three targets have been digitised.

this instance, the 1st, 2nd, 3rd and 4th returns contribute, 80.4 %, 17.0 %, 2.4 % and 0.2 % of the total returns respectively.

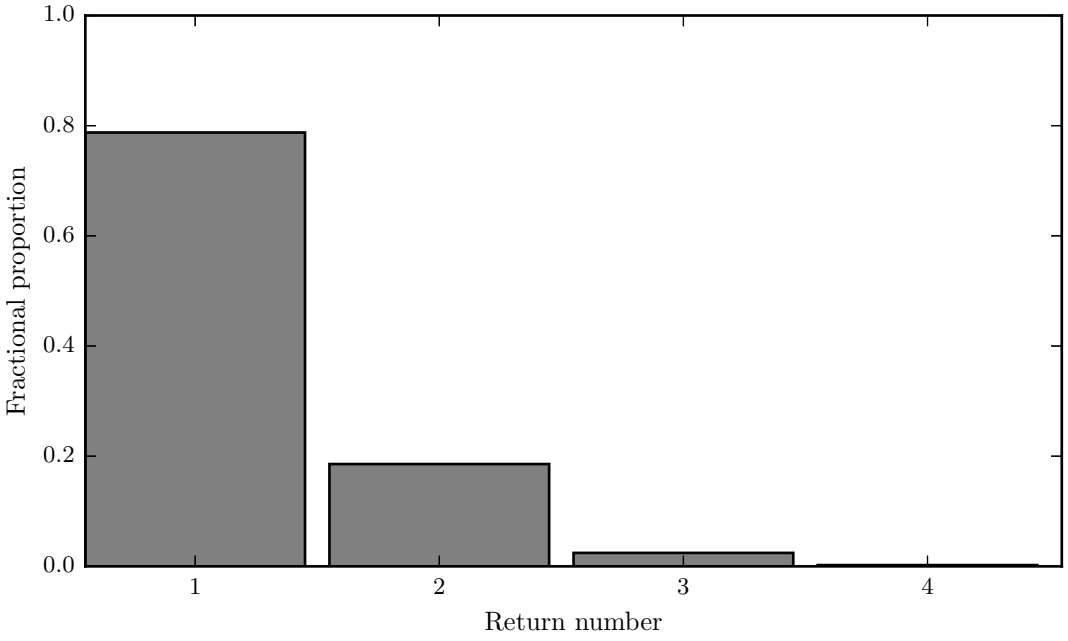


Figure 4.6: Distribution of returns across the point cloud shown in figure 4.2 by return number.

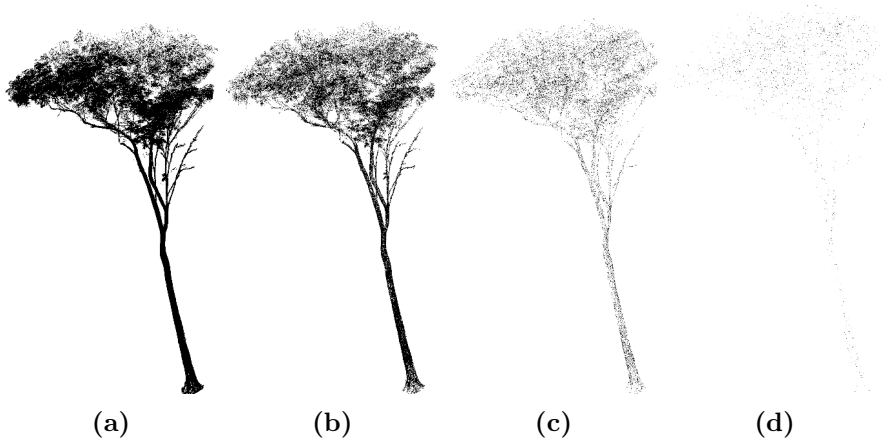


Figure 4.7: Tree-level point cloud representing a dominant canopy tree from tropical forest, segmented into the four point clouds by return number: (a) 1st, (b) 2nd, (c) 3rd and (d) 4th returns.

4.1.3 Radiometric calibration

For each digitised target, a value referred to as 'amplitude', denoted here as A, is provided by the manufacturer. This goes further than the signal amplitude in the traditional sense (i.e., in figure 4.5, subtracting the digital number at the noise floor from the decomposition-inferred digital number at the digitised target). Here, the inferred

digital number is instrument-specifically calibrated to received optical power, P_r, with A expressed in log-space as (Pfennigbauer et al. 2010):

$$A = 10 \log \left(\frac{P_r}{P_{min}}\right) \tag{4.1}$$

Where P_{min} denotes the minimum detectable optical power. A, when referring to the general LiDAR equation in section 3.1.1 equation 3.3, can therefore be equated to the received optical power, P_R.

Again with reference to the general LiDAR equation, explicit characterisation of the beam spread and the system and atmospheric attenuation components, reduces A to a function of the backscatter cross-section. Thus, for each digitised target, an estimate of 'relative reflectance', ρ_{rel}, is derived, defined as the ratio between A and the amplitude from a diffuse, 100 % reflective target at the same range, A_{ref}, such that (Pfennigbauer et al. 2010):

$$\rho_{rel} = A - A_{ref} \tag{4.2}$$

Where A_{ref} is interpolated from an instrument-specific lookup table generated at the manufacturer's test facility.

In an idealised scenario where the pulse footprint fully intercepts a diffuse target oriented normal to the scanner, the target reflectance, ρ, at 1550 nm, can be estimated as:

$$\rho = \left[\frac{e^{\rho_{rel}}}{10}\right]^{-1} \tag{4.3}$$

However, in field conditions these assumptions are rarely, if ever, going to be met, making interpretation of ρ_{rel} impractical (e.g., to discriminate between returns from leaf and wood). This is demonstrated in figure 4.8, where the distribution of ρ_{rel} across all returns inside the point cloud from the single scan acquired in tropical forest from figure 4.2 is shown.

The distribution appears somewhat Gaussian with a heavy negative skew; there is no evidence of bimodality that would allow naive classification of returns by material type, such as leaf and wood. This would be expected, as with reference to the backscatter

cross-section in equation 3.3, insurmountable ambiguities exist because ρ_{rel} is also a function of the angle of incidence between the instrument and target, the directionality of the scattering and the proportion of the beam footprint that intercepts the target. This problem deepens when considering multiple returns, as by definition they do not conform to these assumptions.

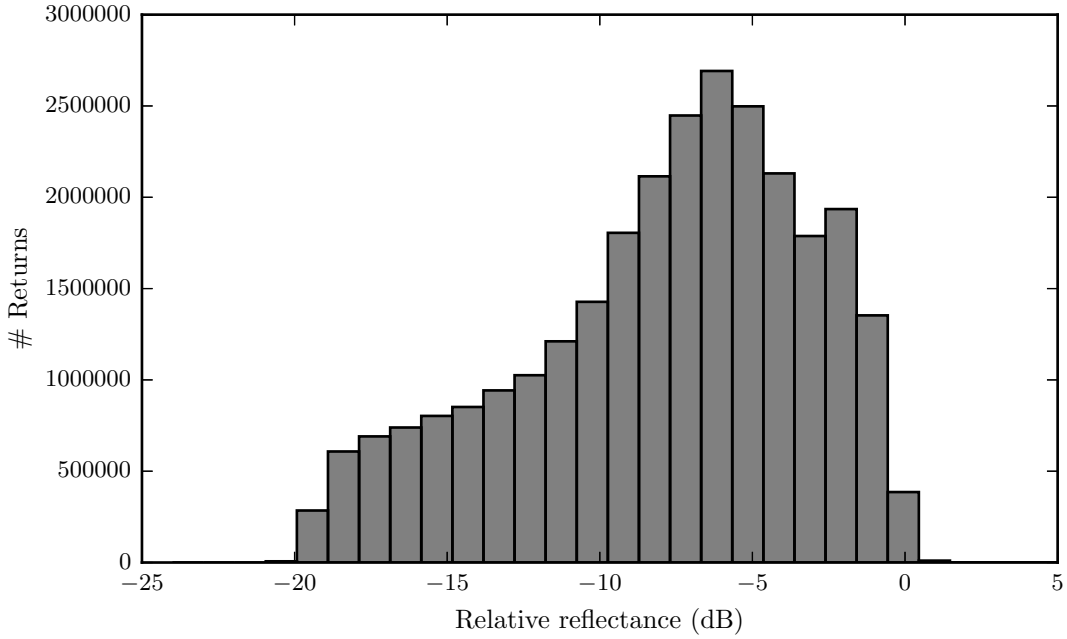


Figure 4.8: Distribution of returns across the point cloud shown in figure 4.2 by ρ_{rel}.

4.1.4 Scan co-registration

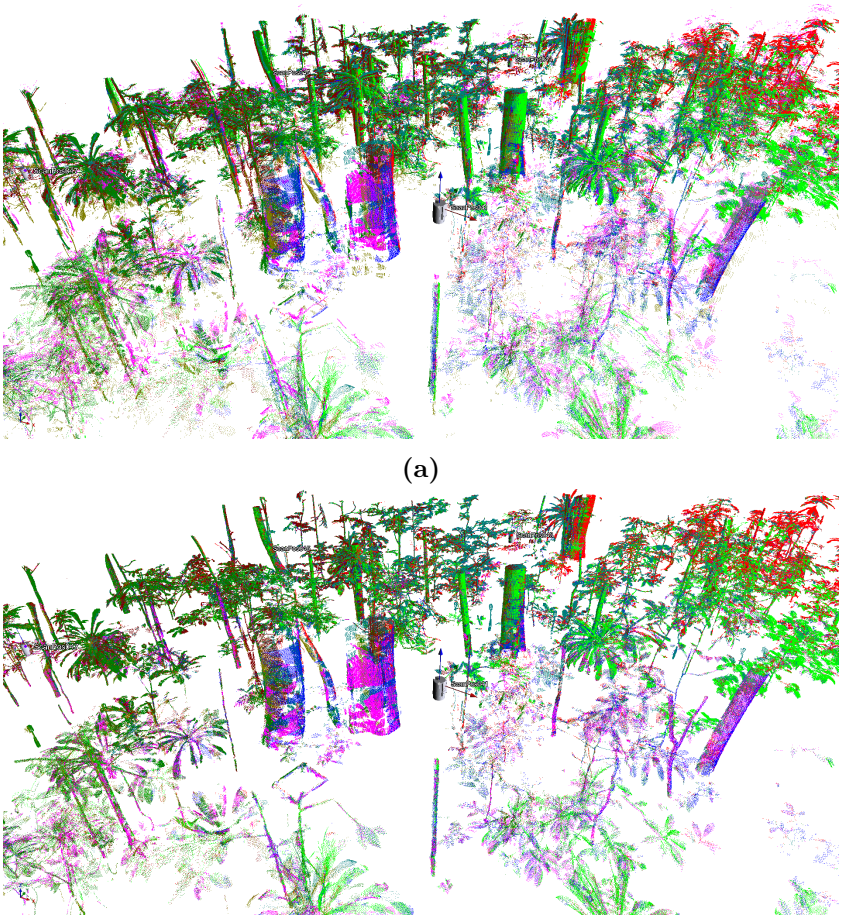
Individual scans are mapped onto a common coordinate system to generate the plot-level point cloud through a two-stage process. In the instance of data acquisition using the 10 m resolution systematic square grid sampling pattern, the first and second scans are located at plot coordinates (0,0) and (0,10) respectively. Retroreflective cylindrical targets are positioned between these two scan locations such that four or more of these targets are observed from both locations. During post-processing, discrimination by ρ_{rel}, allows automatic identification of these targets. Co-registration of the second scan into the coordinate system of the first scan is then achieved by matching the retroreflective targets in each scan, through correspondences estimation, and rotating/translating the second point cloud about this solution. A 'daisy chain' approach is taken to link the third scan (0,20) to the second scan, the fourth scan (0,30) to the third scan, and so on.

This process is demonstrated in figure 4.9a, where a slice of plot-level point cloud

has been constructed from 9 uniquely coloured single scan point clouds acquired from tropical forest in Lopé National Park, Gabon. The accuracy of the co-registration using this first stage is usually at the centimetre level (i.e., an order of magnitude larger than the ranging accuracy). This poor accuracy is either due to incomplete sampling, or clumping, of the retroreflective targets. In dense understory, the targets will often be tightly positioned to each other for ease. This restricts their possible placement to a very narrow corridor through the scanner field of view, increasing the degree of freedom in the correspondences estimation solution, particularly in the rotation. In addition to this, the 'daisy chain' style ensures any errors in the co-registration of a scan will propagate fully into the co-registration of subsequent scans.

The second stage improves the accuracy of the co-registration through mass plane fitting to the single scan point clouds. Plane fitting is defined in chapter 5, but here, at the most simple, involves fitting a plane model through small subsets of the single scan point clouds. Generated planes are considered valid, provided they adhere to user-defined tolerances, such as the error terms (e.g., the average distance between the points and plane). With the first stage providing a coarse plot-level point cloud, it is trivial to identify common planes between single scan point clouds using metrics such as the distance and angle between planes. Identified common planes have the effect of adding a significant number of additional tie points for the correspondences estimation, from across the full extent of the point clouds, improving the accuracy of the co-registration down to the millimetre level. This is shown in figure 4.9b, where the second stage of the co-registration has been applied to the example of figure 4.9a.

Using these methods, the accuracy of the co-registration is the same as the ranging accuracy (5 mm). That is, individual scans are visually indistinguishable from one another in the plot-level point cloud, as can be observed in figure 4.9b. The full approach described here is achieved using the manufacturer's software, RiSCAN Pro (RIEGL Laser Measurement Systems GmbH 2016).



(a)

(b)

Figure 4.9: Slice of plot-level point cloud, constructed through the co-registration of 9 uniquely coloured single scan point clouds acquired in tropical forest from Lopé National Park, Gabon, using a two-stage registration process: a) a course registration is achieved through the placement and matching of retroreflective targets, and secondly b) the accuracy of the registration is improved through mass plane fitting and matching, such that the individual scans are visually indistinguishable from one another.

4.1.5 Data acquisition protocol

As described in section 3.1.5, there are two user-defined variables that are the primary drivers of data quality. Here, data quality is defined as the completeness to which the derived plot-level point cloud captures the 3D structure of the forest scene, within the bounds of the fixed characteristics of the instrument (i.e., ranging accuracy and beam divergence).

Firstly, the angular resolution, θ, determines the angular step between sequentially fired pulses in both azimuth and zenith. Section 3.1.5 equation 3.7 defined the distance, D_a, between the location of two fired pulses along the same zenith ring towards an extended target at range R. Typical definition of θ would be 0.02°, 0.04° and 0.06°. At

a range of 60 m (a typical range between the instrument and top of a 40 m tropical forest canopy), this would correspond to a D_a value of -7 mm, 14 mm, 35 mm respectively. In this idealised scenario then, this defines that 0.02° will oversample the scene, whilst 0.04° and 0.06° will undersample the scene. However, this is significantly harder to infer in complex 3D scenes, where the distribution of target ranges will be highly variable.

Precision actuation of the head and mirror in the RIEGL VZ-400 permit angular resolutions as low as 0.0024°. The main factor in preventing a definition of θ this low is scan duration. Halving θ results in the emission of four times the number of pulses, and hence, a four time greater scan duration. At 0.04°, a single scan takes approximately 3 min.

The second user-defined variable affecting data quality is the sampling pattern (i.e., the location and number of scans through the plot). Two sampling patterns have been used to collect the data considered in this thesis. The first sampling pattern is a systematic square grid, of 10 m or 20 m resolution cast across the plot boundary. Across a 1 ha forest plot, this results in 121 and 36 scan locations respectively. The second sampling pattern, considered only for 0.25 ha forest plots, uses 4 scan locations along the plot perimeter, on the principal axes, with a single scan at the centre of the plot. The sampling pattern employed to sample a forest plot is a response to the forest plot itself, primarily a function of basal area and the complexity of both the understory and crown.

At each of these scan locations that TLS data were collected, two separate scans were conducted. This is due to the 100° zenith field of view of the RIEGL VZ-400 that prevents the full hemisphere from being sampled in a single scan. For a complete sample then, the scanner must be tilted through 90°. This results in some sampling overlap ($\approx 20°$), although the data would not be considered redundant due to the small optical centre offset induced in all three axes by the action of tilting. The orientation of this tilt was consistent through data acquisition, and remained such that the titled scan would capture the retroreflective targets for co-registration.

Across the five forest plots at which data have been collected for this thesis, the parameters of angular resolution and sampling pattern vary significantly, as illustrated in section 4.2. This variation is mainly a result of the novelty of these measurements, leading to their definition on an ad-hoc basis. In particular, the tropical forest data collected during the work for thesis is the first of its kind, with no previous work available

to influence the data acquisition protocol. As such, the sampling protocol has been consecutively refined between field campaigns.

4.1.6 Noise reduction

Section 4.1.3 defined the deviation metric that accompanies each resolved target as the resemblance between the returning waveform from which the target was digitised and the matched waveform that it was decomposed against from the instrument-specific database. Figure 4.4 presented five waveforms whose deviation metric ranged from 0 to 100 The metric exists to act as a proxy for the goodness of the ranging estimate. By default, the instrument adds any target that is decomposable from the waveform, to the point cloud. Figure 4.10a presents a close-up of a small segmented section of the point cloud presented in figure 4.2. It can be seen that a number of 'ghost points' exist around seemingly valid returns from both clumps of leafy material and returns from the edges of stems.

This is a result of the waveform decomposition methods being unable to satisfactorily resolve noisy data. Their impact on the final point cloud is removed by filtering via this deviation metric. The manufacturer recommends that deviation values between the range of $10 - 20$ and above should be discarded. Through trial and error, it was decided here that returns with deviation values exceeding 15 are pruned from the final single scan point clouds. Figure 4.10b shows the same section of point cloud after this filtering has been implemented. It can be seen that the vast majority of the 'ghost points' are removed, whilst having a minimal impact on valid returns.

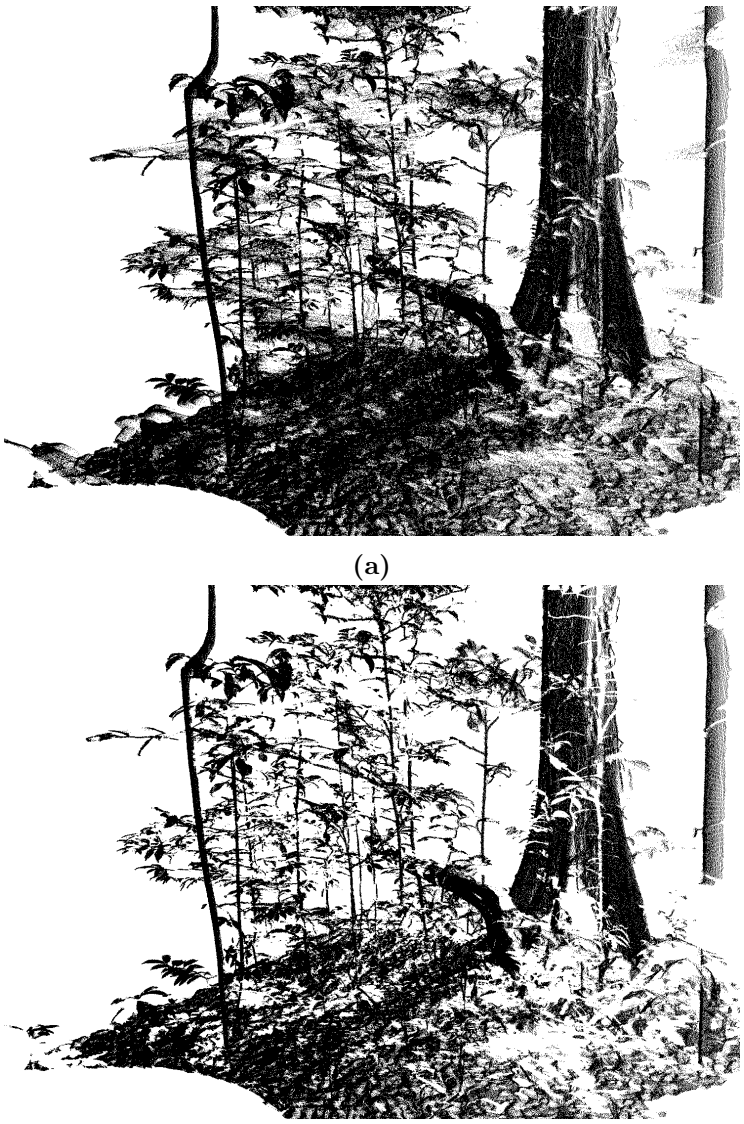


Figure 4.10: Section of point cloud segmented from figure 4.2, where (a) the point cloud comprises all digitised targets, and (b) the point cloud has been filtered by the deviation metric, to remove the 'ghost points' visible in (a) around clumps of leafy material and the edges of stems.

4.1.7 Environmental influences on the data

Any error in the sampling of the scene is likely to arise from two environmental factors. The first source is wind, disturbing the structural elements of the forest, particularly in the canopy, such that they will be observed in different positions or arrangements across scans. In the mild form, this will manifest as an increase in the apparent size of branching or leafy clumps, as shown in figure 4.11. Or, more severely, the structure will be sufficiently displaced from rest that it is duplicated in the plot-level point cloud, as also shown in figure 4.11. The influence of wind on the quality of the plot-level point

cloud is reduced by scanning in low wind speed conditions. This is not always practical within the time constraints of a field campaign, so a trade-off between acquisition and contamination is often made. All data considered in this thesis is assumed to be affected by wind, although in some, even many cases, this is not true. Discretion was used to cease scanning where necessary, in the context of the objective of retrieving tree-scale AGB from the data.

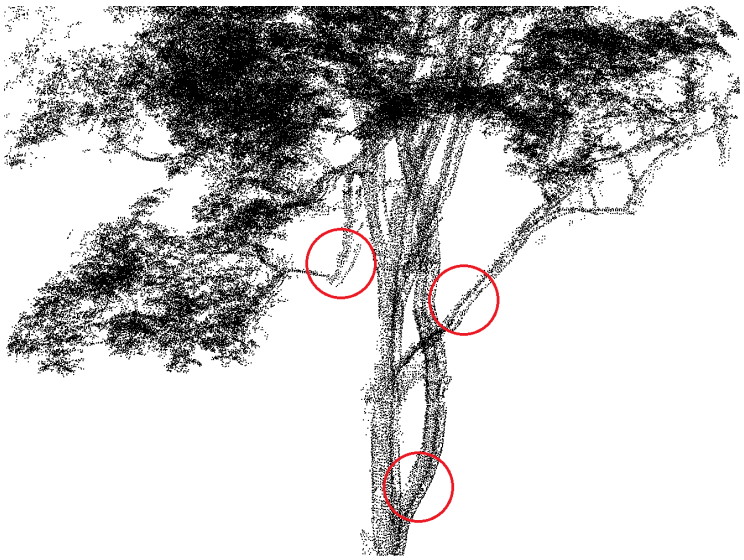


Figure 4.11: Tree-level point cloud representing a dominant canopy tree, segmented from the plot-level point cloud acquired in Nouragues Nature Reserve, French Guiana; influence of wind on the point cloud is highlighted in three areas, the lower area identifies the displacement of branching between scans, giving rise to an apparent increase in the size of the branching, whilst the other two areas show more severe contamination where the branching appears duplicated.

The second environmental influence is precipitation. Heavy rainfall can also disturb the structural elements of the forest, particularly the position and arrangement of leafy material; this will again manifest as an apparent increase in the size of clumps of leafy material. In addition to this, as shown in figure 3.2, water absorbs the significant proportion of radiation at 1550 nm, resulting in rainfall deposition on forest elements influencing target derived ρ_{rel}. The contamination of the data considered in this thesis by precipitation is reasonably small as the field campaign were scheduled during dry seasons. In the instances where significant rainfall occurred, scanning was ceased as the instrument is only minimally weatherproof. A reasonable amount of time was then taken before recommencing data acquisition to allow the scene to dry somewhat.

4.2 Overview of the field sites

The geographical locations from which TLS data have been collected during the work for this thesis are shown in figure 4.12. TLS data have been acquired from 3 ha of censused African and South American moist, Terra Firme, lowland, mixed species, old-growth tropical forest. With deployment of a TLS instrument never before undertaken in tropical forest, these data are the first of their kind. Data has also been collected from 0.25 ha of censused *Eucalyptus* spp. open forest in Australia. In addition to this, 3 *Quercus* spp. individuals from the UK have been destructively harvested (with direct measurement of AGB made) post-scanning (in leaf-on and off conditions); to provide a unique validation dataset.

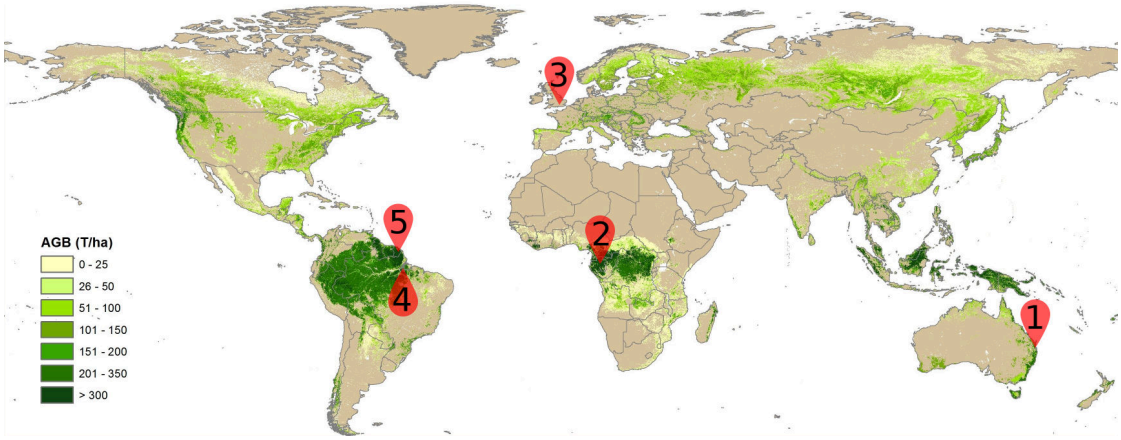


Figure 4.12: Location of the five forest plots from which TLS data has been acquired during the work for this thesis, overlain a global AGB map (Source: Avitabile et al. 2016).

The following sections individually review the five forest plots. Each section begins with a set of wide-angle photographs acquired from a Nikon D300 camera mounted on the RIEGL VZ-400. This is followed by a table containing: a description of the forest type and location, a definition of the census data, the quantification of relevant structural parameters and a detailing of the scanning acquisition protocol. A stem map is then provided for each forest plot overlaid on the adopted sampling pattern. Finally, a slice of the plot-level point cloud is then shown.

4.2.1 Karawatha Forest Park, Australia (KARA-001)

TLS data were acquired from a single 0.25 ha censused *Eucalyptus* spp. open forest plot in Karawatha Forest Park, Australia (designation: KARA-001) in July 2013. Figure 4.13 presents five wide-angle photographs taken in the plot. The characteristics of the plot are provided in table 4.2.

KARA-001 is the least structurally complex plot of those considered here. In essence, the forest was arranged in a single canopy layer with a canopy height of approximately 25 m, although there was some presence of younger trees sparsely located through understory. The undergrowth consisted of grasses (less than 0.5 m tall) occasionally interrupted by small shrubs. With a stem count of 40 and a basal area of 10.43 $m^2\,ha^{-1}$, the plot was lightly populated. As a result, it can be seen in figure 4.13 there were large gaps present in the canopy with little to no crown interaction. This permitted KARA-001 to be sufficiently sampled using a 0.06° angular resolution and a 5 scan principal axes sampling pattern, as illustrated in figure 4.14, overlaid on the plot stem map.

Using the NLS allometric model described in chapter 2, with allometric regressors D and ρ, plot-scale biomass was estimated at 36 663 kg, with a relative uncertainty of 53.5 %. Per unit area, this makes KARA-001 the lowest AGB forest plot sampled. A slice of the acquired plot-level point cloud, is shown in figure 4.15.

Figure 4.13: Wide-angle photographs from KARA-001, a *Eucalyptus* spp. open forest plot in Karawatha Forest Park.

Plot code	KARA-001
Location	Karawatha Forest Park
Country	Australia
Longitude	$-27.63°$
Latitude	$153.07°$
Forest type	*Eucalyptus* spp. open forest
Plot size	0.25 ha
Tree-scale parameters measured in census	x, y, D, ρ
Minimum D considered for census	0.1 m
Stem count	40
Basal area	$10.43\,\mathrm{m^2\,ha^{-1}}$
Canopy height (approximate)	25 m
Basal area-weighted average mean wood density	$840\,\mathrm{kg\,m^{-3}}$
AGB (via $D^2\rho$ NLS allometric)	36 663 kg
AGB lower 95 % interval	27 094 kg
AGB upper 95 % interval	46 700 kg
AGB relative uncertainty	26.8 %
TLS data acquisition date	July 2013
Angular resolution	$0.06°$
Scan pattern	Centre and perimeter principal axes
Scan locations	5

Table 4.2: Parameters of forest plot KARA-001 and description of the TLS sampling regime.

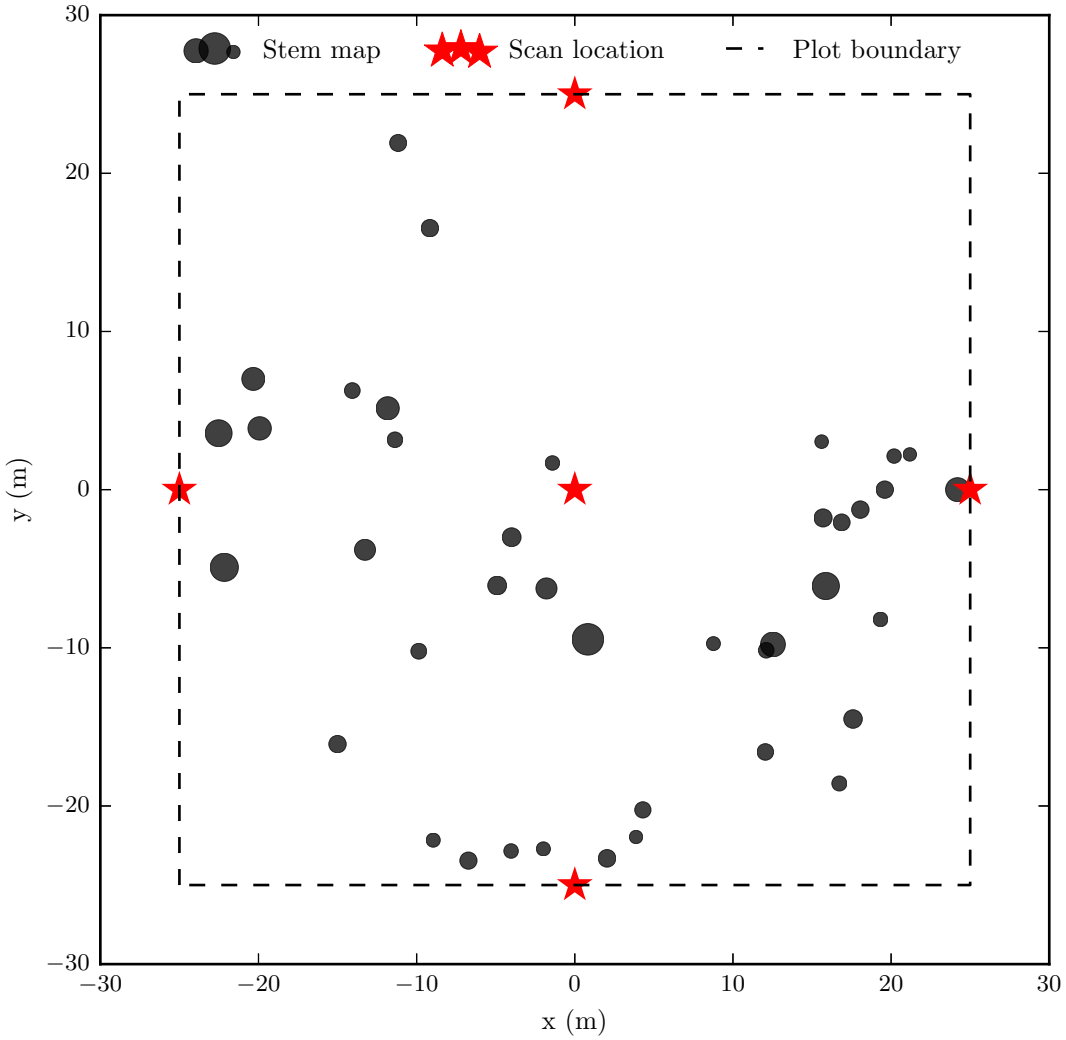


Figure 4.14: Principle axes sampling pattern used for TLS data acquisition in forest plot KARA-001, overlaid on the plot stem map and perimeter.

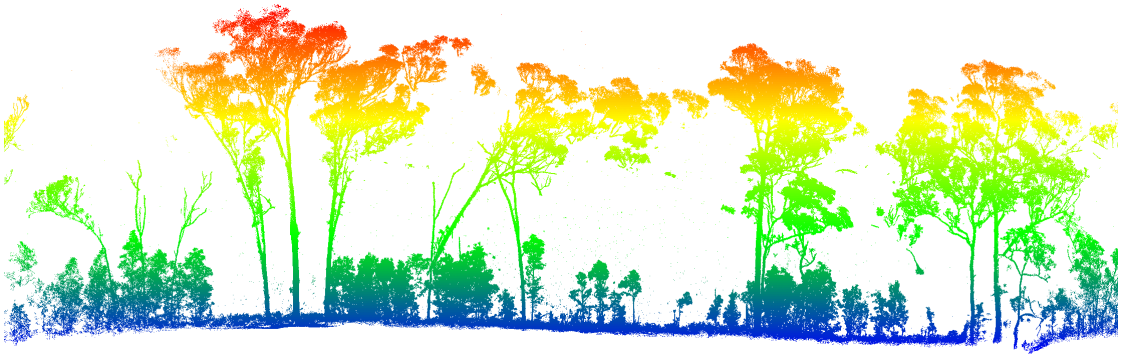


Figure 4.15: Slice of KARA-001 plot-level point cloud, coloured by height.

4.2.2 Lopé National Park, Gabon (LPG-01)

TLS data were acquired from a single 1 ha censused moist, Terra Firme, lowland, mixed species, old-growth tropical forest plot in Lopé National Park, Gabon (designation: LPG-01) in August 2013. Figure 4.16 presents five wide-angle photographs taken in the plot. The characteristics of the plot are provided in table 4.3.

The forest in LPG-01 was multilayered with a canopy height of approximately 40 m. An emergent layer was present with a few individuals of height exceeding 45 m. The understory contained a significant number of young and shade tolerant trees. The undergrowth was relatively sparse although occasionally interrupted by patches of dense mixed vegetation around tree fall.

Of the three tropical forest plots sampled, LPG-01 was the least structurally complex. This was due to stem count of 342 , which is some 44 % lower than NOU-11 (subsection 4.2.5). Despite this, as can be seen in figure 4.16, the gap fraction was relatively low and there was significant interaction between neighbouring crowns across the layers.

This was the first tropical forest from which TLS data were acquired so the sampling regime was somewhat improvised. An angular resolution of 0.06° and a 20 m resolution systematic square grid sampling pattern was adopted for data acquisition, as shown in figure 4.17, overlaid on the plot stem map. This regime provided an adequate sampling of the plot, as seen from the slice of plot-level point cloud presented in figure 4.18.

Using the NLS allometric model developed in chapter 2, with allometric regressors D, H and ρ, AGB is estimated at 470 224 kg. Despite the low stem count, LPG-01 nominally represents the largest AGB per unit area plot considered (although this inference is statistically prohibited by a relative uncertainty in this estimate of 62.2 %). This is because of the presence of several large stems, in particular a single individual (D = 2.51 m) contributing 26.3 % of plot-scale AGB.

Figure 4.16: Wide-angle photographs from LPG-01, a moist, Terra Firme, lowland, mixed species, old-growth tropical forest plot in Lopé National Park.

Plot code	LPG-01
Location	Lopé National Park
Country	Gabon
Longitude	$-0.18°$
Latitude	$11.57°$
Forest type	Tropical
Forest moisture	Moist
Forest edaphic type	Terra Firme
Forest elevation	Lowland
Forest composition	Mixed species
Forest status	Old-growth
Plot size	1 ha
Tree-scale parameters measured in census	x, y, D, H, ρ
Minimum D considered for census	0.1 m
Stem count	342
Basal area	$31.54\,\mathrm{m^2\,ha^{-1}}$
Canopy height (approximate)	40 m
Basal area-weighted average mean wood density	$746\,\mathrm{kg\,m^{-3}}$
AGB (via $D^2H\rho$ NLS allometric)	470 224 kg
AGB lower 95 % interval	344 114 kg
AGB upper 95 % interval	636 596 kg
AGB relative uncertainty	31.1 %
TLS data acquisition date	August 2013
Angular resolution	$0.06°$
Scan pattern	20 m systematic square grid
Scan locations	36

Table 4.3: Parameters of forest plot LPG-01 and description of the TLS sampling regime.

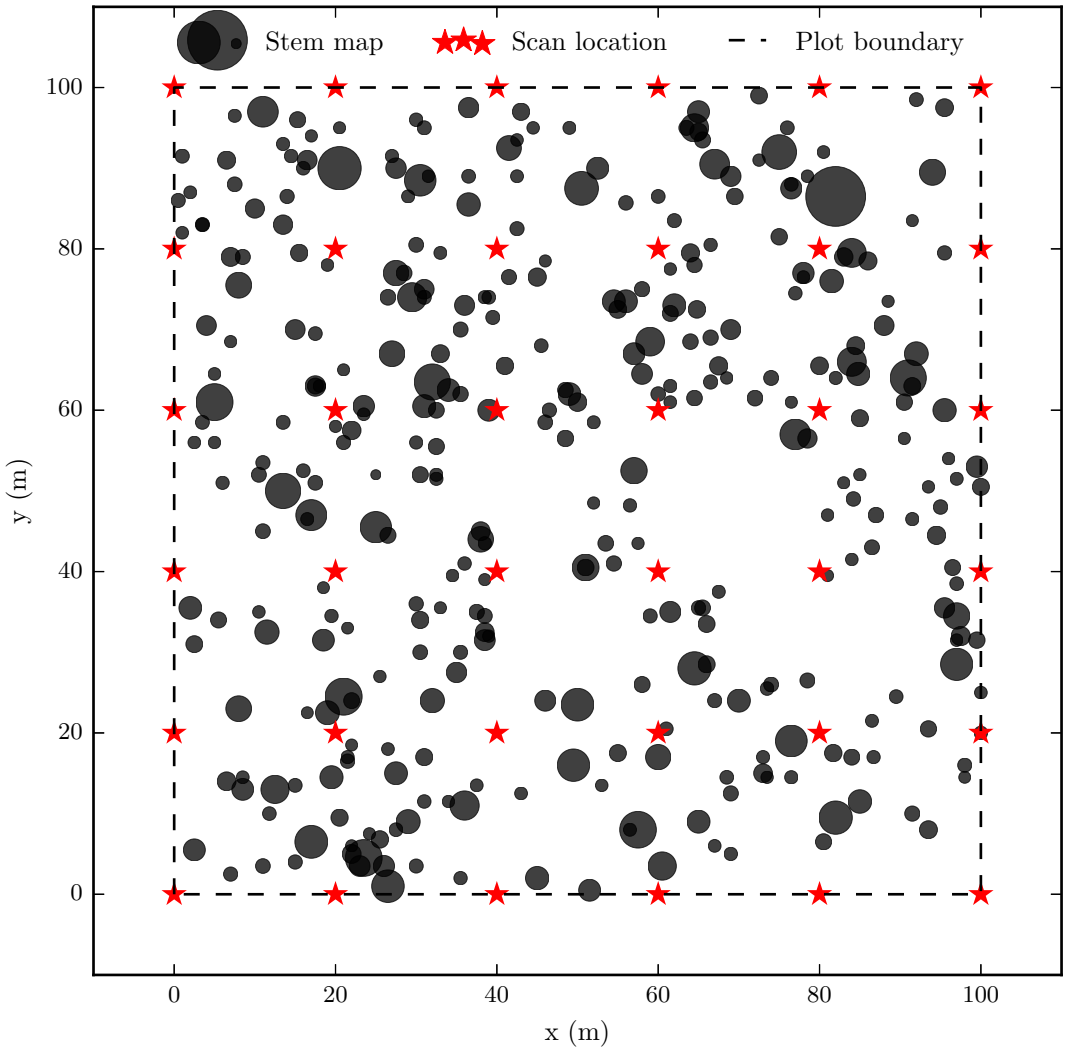


Figure 4.17: 20 m resolution systematic square grid used for TLS data acquisition in forest plot LPG-01, overlaid on the plot stem map and perimeter.

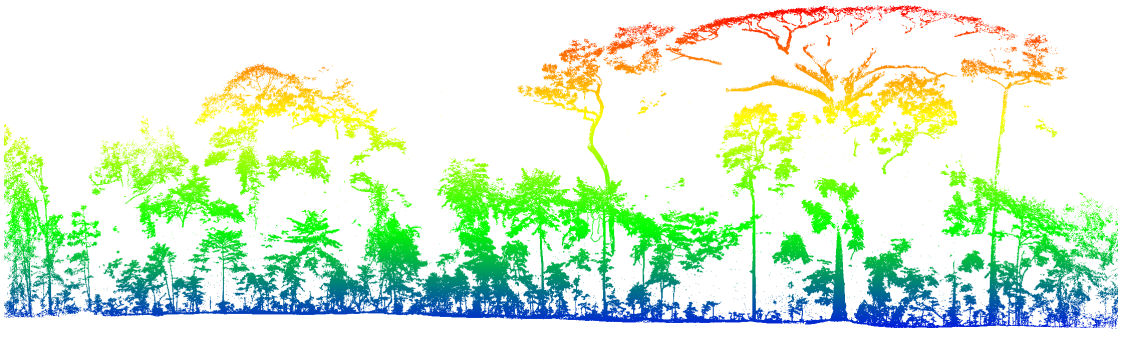


Figure 4.18: Slice of LPG-01 plot-level point cloud, coloured by height.

4.2.3 Alice Holt Forest, UK (AH)

The TLS data collected from *Quercus* spp. woodland in Alice Holt Forest, UK (coordinates: 51.15,-0.85) differs from other data considered in this thesis as the acquisition was focused at the tree-level. The three individuals of interest (designation: AH-01, AH-02 and AH-03) are shown in the portrait photographs of figure 4.19. These three broad-leaved trees were scanned in April 2014 and July 2014 to sample in both leaf-off and leaf-on conditions respectively.

In July 2014, the Forestry Commission harvested the trees and direct measurement of AGB and ρ was made. The results of these destructive harvest measurements are provided in table 4.4. Prior to weighing, the trees were partitioned into wood and leaf; whereby, averaged across the three individuals, wood contributed 97.5 % of total AGB.

As can be seen in figure 4.19, the forest in Alice Holt is relatively sparse and there was no interaction between the trees in question with any neighbours. Additionally, any undergrowth was completely removed prior to scanning. This allowed a simple scanning pattern, with 4 scan locations on the principal axes relative to the tree, at a range of 5 m from the tree, to be used. Using an angular resolution of 0.06°, the tree-level point clouds shown in both leaf-off and leaf-on conditions in figure 4.20, were produced.

Tree-code	AH-01	AH-02	AH-03
Wood AGB (kg)	434	1242	638
Leaf AGB (kg)	13	29	17
Total AGB (kg)	447	1271	655
Average wood density ($\mathrm{kg\,m^{-3}}$)	592	535	531

Table 4.4: Destructive harvest measurements of AH-01, AH-02 and AH-03.

Figure 4.19: Wide-angle photographs of (top row left to right) AH-01, AH-02 and AH-03 from *Quercus* spp. woodland in Alice Holt Forest.

Figure 4.20: Tree-level point clouds of (left to right) AH-01, AH-02 and AH-03 in leaf-off (top row) and leaf-on (bottom row) conditions.

4.2.4 Caixuanã National Forest, Brazil (CAX-A)

TLS data were acquired from a single 1 ha censused moist, Terra Firme, lowland, mixed species, old-growth tropical forest plot in Caixuanã National Forest, Brazil (designation: CAX-A) in October 2014. Figure 4.21 presents five wide-angle photographs taken in the plot. The characteristics of the plot are provided in table 4.5.

Similar to LPG-01, the forest in CAX-A was multilayered with a canopy height of approximately 41 m, with the maximum tree height in the emergent layer of 55 m. The understory contained a mixture of young and shade tolerant trees. In contrast to LPG-01, there was a significant undergrowth component, with dense mixed vegetation through the first few meters off the forest floor, that, in areas was impassable. Overall, in comparison to LPG-01, CAX-A would be considered a more structurally complex forest plot. A divergent stem count but comparable AGB and basal area results in a more uniform distribution of AGB across a larger number of stems. This increased stem count contributed to the significant interaction between neighbouring crowns across the layers.

Due to the structural complexity of the plot, a 20 m resolution systematic square grid and an increased angular resolution of 0.04° was used. This sampling regime is shown in figure 4.22, overlaid on the plot stem map. A slice of plot-level point cloud produced from this sampling regime is shown in figure 4.23.

Figure 4.21: Wide-angle photographs from CAX-A, a moist, Terra Firme, lowland, mixed species, old-growth tropical forest plot in Caixuanã National Forest.

Plot code	CAX-A
Location	Caixuanã National Forest
Country	Brazil
Longitude	$-1.74°$
Latitude	$-51.46°$
Forest type	Tropical
Forest moisture	Moist
Forest edaphic type	Terra Firme
Forest elevation	Lowland
Forest composition	Mixed species
Forest status	Old-growth
Plot size	1 ha
Tree-scale parameters measured in census	x, y, D, ρ
Minimum D considered for census	0.1 m
Stem count	445
Basal area	$33.24\,\mathrm{m^2\,ha^{-1}}$
Canopy height (approximate)	41 m
Basal area-weighted average mean wood density	$710\,\mathrm{kg\,m^{-3}}$
AGB (via $D^2\rho$ NLS allometric)	455 953 kg
AGB lower 95 % interval	347 560 kg
AGB upper 95 % interval	578 665 kg
AGB relative uncertainty	25.4 %
TLS data acquisition data	October 2014
Angular resolution	$0.04°$
Scan pattern	20 m systematic square grid
Scan locations	36

Table 4.5: Parameters of forest plot CAX-A and description of the TLS sampling regime.

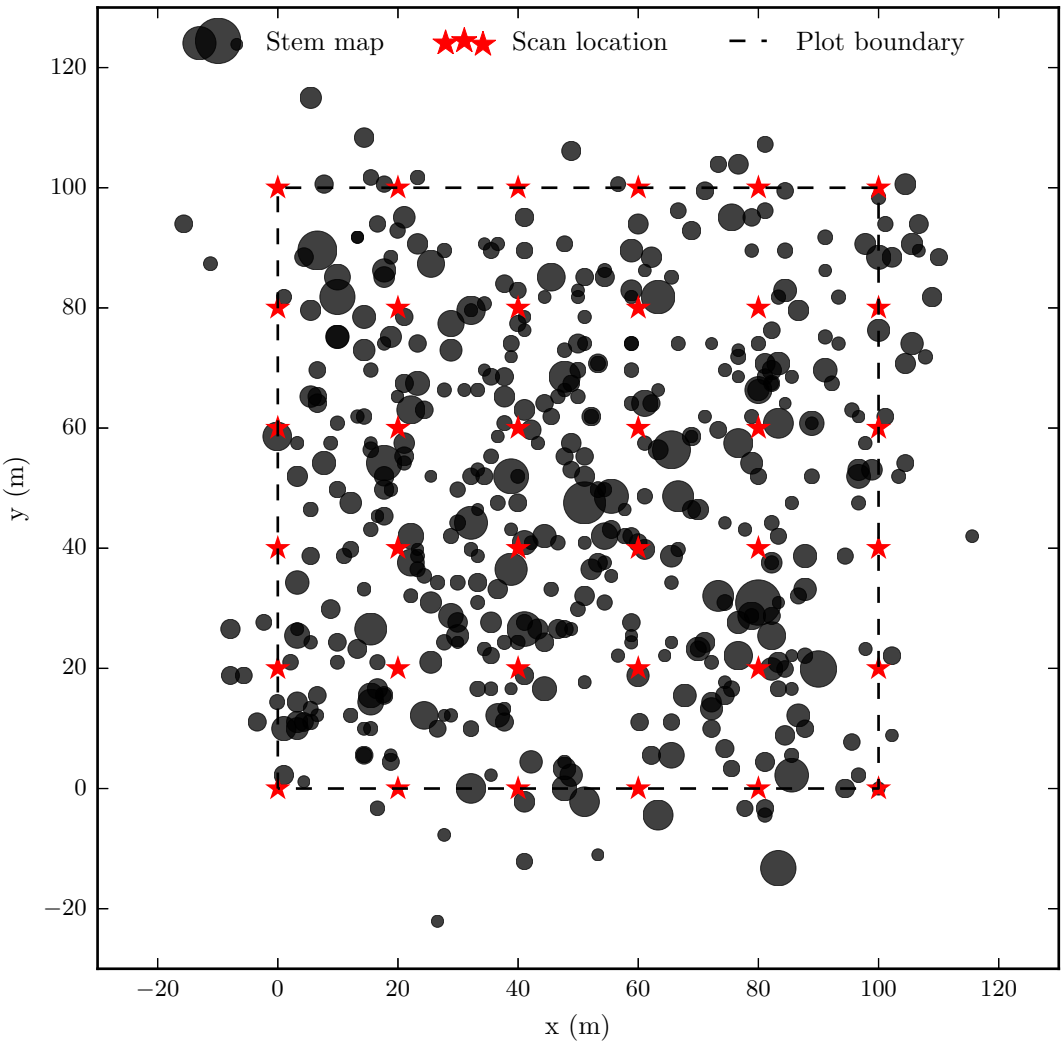


Figure 4.22: 20 m resolution systematic square grid used for TLS data acquisition in forest plot CAX-A, overlaid on the plot stem map and perimeter.

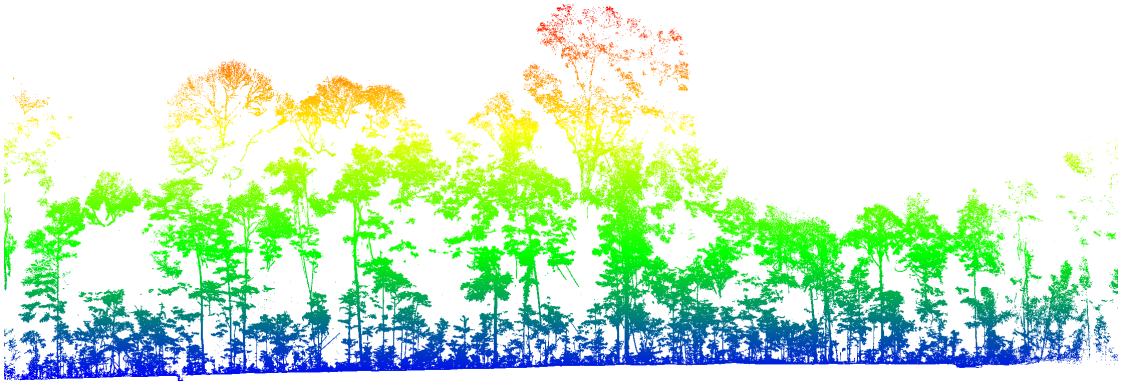


Figure 4.23: Slice of the CAX-A plot-level point cloud, coloured by height.

4.2.5 Nouragues Nature Reserve, French Guiana (NOU-11)

TLS data were acquired from a single 1 ha censused moist, Terra Firme, lowland, mixed species, old-growth tropical forest plot in Nouragues Nature Reserve, French Guiana (designation: NOU-11) in November 2015. Figure 4.24 presents five wide-angle photographs taken in the plot. The characteristics of the plot are provided in table 4.6.

As with the other tropical forest plots, the forest in NOU-11 was multilayered with a canopy height of approximately 41 m, with a maximum tree height in the emergent layer of 52 m. The understory layer contained a mixture of young and shade tolerant trees; in particular, a fairly uniform and dense distribution of trees from the *Arecaceae* family. Their large area leaves, oriented horizontal to the forest floor considerably reduce the gap fraction in the images from figure 4.24. The undergrowth was also significant by variable across the plot, consisting of dense mixed vegetation. Finally, there was some liana encroachment in pockets across the plot.

Of the three tropical forest plots sampled, subjectively, NOU-11 was the most structurally complex. Despite a nominally lower AGB than both LPG-01 and CAX-A (when uncertainties are ignored), a large amount of vegetation is present with a stem count of 425 and basal area of 33.13 $m^2\,ha^{-1}$. This ensured there was significant interaction between neighbouring crowns across the layers.

Arising from this structural complexity and the anticipated understory driven occlusion, a 10 m systematic square grid was used with 0.04° angular resolution, as shown in figure 4.25, overlaid on the plot stem map. A slice of the plot-level point cloud is shown in figure 4.26.

Figure 4.24: Wide-angle photographs from NOU-11, a moist, Terra Firme, lowland, mixed species, old-growth tropical forest plot in Nouragues Nature Reserve.

Plot code	NOU-11
Location	Nouragues Nature Reserve
Country	French Guiana
Longitude	$-52.68°$
Latitude	$4.08°$
Forest type	Tropical
Forest moisture	Moist
Forest edaphic type	Terra Firme
Forest elevation	Lowland
Forest composition	Mixed species
Forest status	Old-growth
Plot size	1 ha
Tree-scale parameters measured in census	x, y, D, H, ρ
Minimum D considered for census	0.1 m
Stem count	425
Basal area	$33.13\,\mathrm{m^2\,ha^{-1}}$
Canopy height (approximate)	45 m
Basal area-weighted average mean wood density	$690\,\mathrm{kg\,m^{-3}}$
AGB (via $D^2H\rho$ NLS allometric)	439 975 kg
AGB lower 95 % interval	339 621 kg
AGB upper 95 % interval	545 610 kg
AGB relative uncertainty	23.4 %
TLS data acquisition date	November 2015
Angular resolution	$0.04°$
Scan pattern	10 m systematic square grid
Scan locations	121

Table 4.6: Parameters of forest plot NOU-11 and description of the TLS sampling regime.

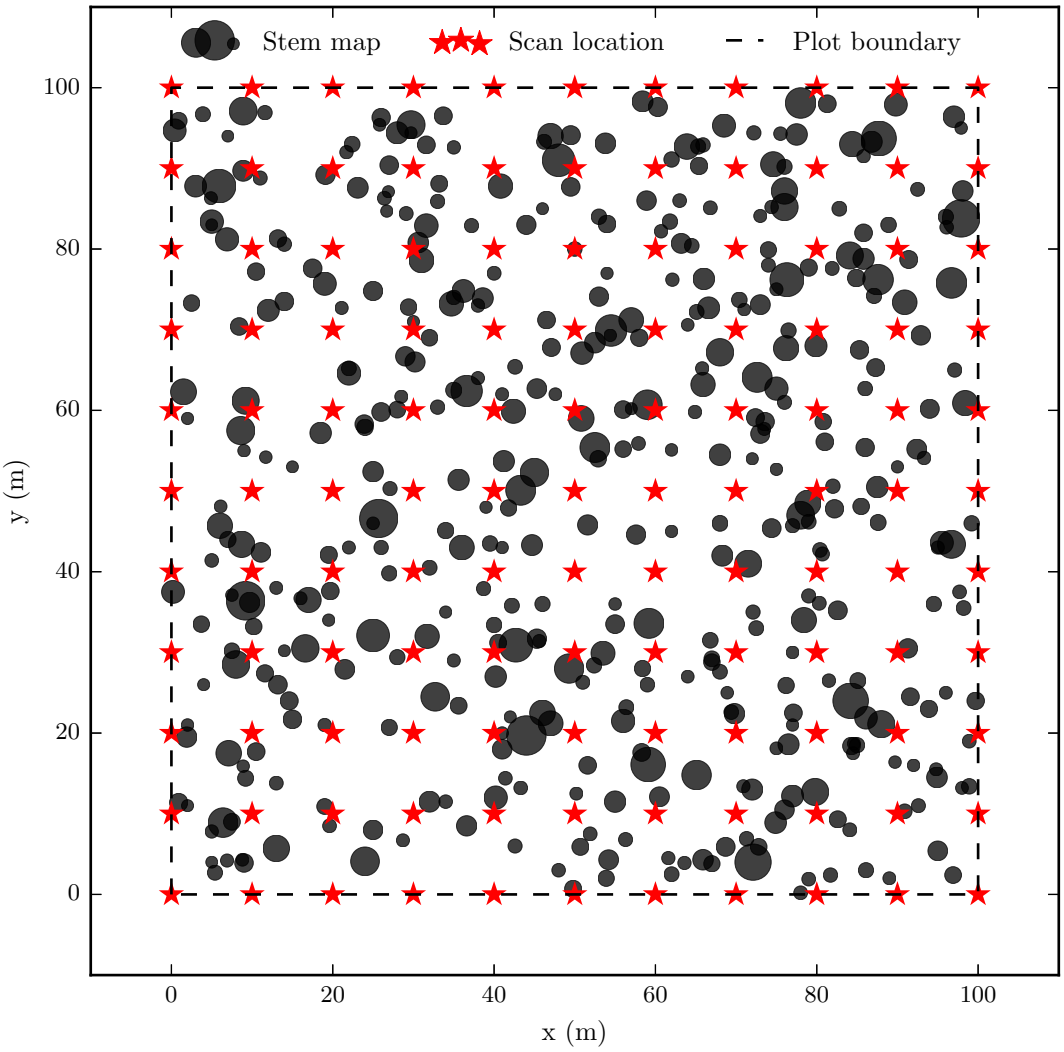


Figure 4.25: 10 m resolution systematic square grid used for data acquisition in forest plot NOU-11, overlaid on the plot stem map and perimeter.

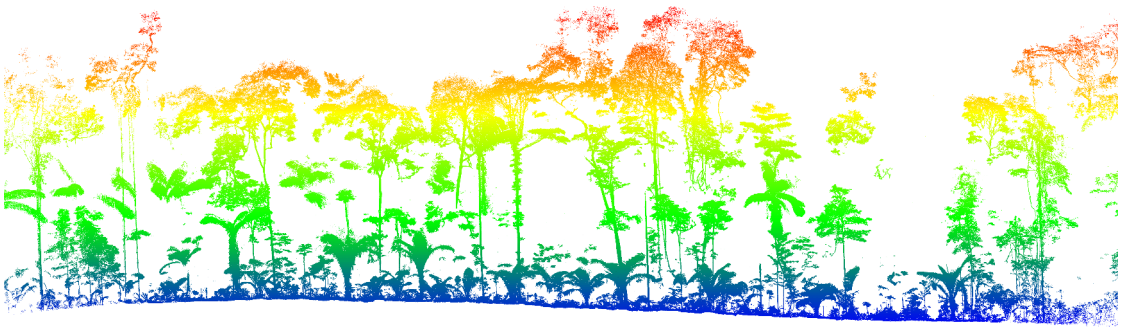


Figure 4.26: Slice of NOU-11 plot-level point cloud, coloured by height.

5 *treeseg*: an algorithm for extracting tree-level point clouds

5.1 Introduction

Chapter 2 demonstrated through a novel experimental study that across large swathes of tropical forest, allometric-derived estimates of AGB are statistically indistinguishable from one another because of large uncertainties that exceed 75 % and 25 % of AGB at the tree- and plot-scale respectively. This study illustrated that there was a requirement for the introduction of new, more accurate methods for the non-destructive estimation of tree- and plot-scale AGB.

Chapter 3 provided a general review of terrestrial laser scanning (TLS), the alternative method considered in this thesis to estimate tree- and plot-scale AGB. It was shown that commercial off-the-shelf TLS instruments are capable of rapidly and accurately obtaining a near-complete 3D sampling of a forest scene that is realised through the construction of a plot-level point cloud. A number of studies were then reviewed whose methods fitted geometric primitives to tree-level point clouds to reconstruct tree surface topography. These quantitative structural models (QSMs), particularly those constructed using the methods of Raumonen et al. (2013) and Hackenberg et al. (2015), were shown to be capable of explicitly characterising woody tree structure.

Using these QSMs, an estimate of tree-scale volume can be derived from the summed volume of each geometric primitive. Alongside a value of wood density then, an estimate of tree-scale AGB can be retrieved via volume estimation. By using these methods there is no requirement to infer the influence of tree structure on AGB, as would be necessary in allometry, as tree structure has been measured at source. These methods then, are entirely independent of allometry and the large uncertainties demonstrated in chapter 2 that burden allometric methods. Proof of concept of these new methods has been provided by the studies of Disney et al. (2012) and Calders et al. (2015) who demonstrated that tree-scale volume and AGB could be accurately retrieved from simulated and real validation datasets. As a result, these new methods present a potential paradigm shift in the way that tree-scale AGB is measured.

Chapter 3 concluded by stating the three knowledge gaps that currently prohibit

application of these new methods on a wide-scale (here meaning application across tens of 1 ha forest plots). The implementation of these new methods on a wide-scale is a fundamental requirement for them to become a viable alternative to allometric methods.

The first of these knowledge gaps is the automatic extraction of tree-level point clouds from the plot-level point cloud. The literature reviewed in chapter 3, whose objectives aimed towards the retrieval of tree-scale structure from TLS data, near-ubiquitously extracted tree-level point clouds using 'by hand extraction'. That is, plot-level point clouds were manually manipulated by the user to extract individual tree-level point clouds. Across a number of forest plots this method is impractical; this impracticality is compounded still further when complex forest types are considered. For instance, in the tropical forest plots from which TLS data have been acquired during the work for this thesis, the complexity arising from vegetation interaction would make 'by hand extraction' inconceivable.

To date, the sole publication looking to introduce automatic methods is the study of Raumonen et al. (2015). Here, the authors reconstructed a full *Quercus* spp. oak plot in the UK. These data however, were obtained in a relatively sparsely packed plot in leaf-off conditions, such that there was minimal interaction between stems. That is, there have been no publications looking to introduce methods to automate, or even partially automate, the extraction of tree-level point clouds from plot-level point clouds that have been acquired through complex forest scenes.

This chapter describes *treeseg*, a novel algorithm that has been developed during the work for this thesis to bridge this knowledge gap. This algorithm ingests a plot-level point cloud and semi-automatically extracts tree-level point clouds. The resultant tree-level point clouds are either extracted in their entirety from the plot-level point cloud, or in instances of ambiguity, are extracted such that some manual user refinement is then required. As a result, this algorithm overcomes much of the current impracticality associated with accessing the tree-level information content of TLS data. An example of a tree-level point cloud extracted from a section of plot-level point cloud using the *treeseg* algorithm can be seen in figure 5.1.

This chapter is broken into the following sections. Section 5.1.1 describes the underlying design philosophy that drove the development of the algorithm. Next, section 5.2 introduces a number of generic techniques that permit access to the information content

Figure 5.1: Extraction of a tree-level point cloud (red) from a section of plot-level point cloud (black) using the *treeseg* algorithm.

of a point cloud that have been extensively used in the development of *treeseg*. Section 5.3 then describes the methods of the algorithm. Following this, section 5.4 presents the results from the application of *treeseg* to the TLS data acquired from 3.25 ha of predominantly tropical forest, as described in chapter 4 section 4.2. Section 5.4 then discusses these results, and the conclusions of this chapter are presented in section 5.5.

5.1.1 Design philosophy

The development of *treeseg* has been constrained by attempting to adhere as closely as possible to the following three principles:

- **Automation**
- **Independent of forest type**
- **Independent of choice of TLS instrument**

As described in section 5.1, for TLS-derived estimates of tree- and plot-scale AGB to become viable alternatives to allometric modelling it must be practical to implement them on a wide-scale. A core enabler of this will be method-wide automation; i.e., requiring zero, or minimal user input and time required to extract tree-level point clouds from plot-level points clouds obtained across tens of 1 ha forest plots.

Furthermore, these efforts to automatically extract individual tree-level point clouds from plot-level point clouds need to be independent of forest type, as far as possible. That is, algorithm performance should remain constant regardless of the complexity of the forest structure. To aid in achieving this, *treeseg* has been developed using TLS data acquired from tropical forests as these represent some of the most structurally complex encountered (Richards 1952; Clark et al. 2000). Combined with the imposition of little *a priori* expectation of tree architecture, algorithm performance should be readily transferable to less structurally complex forest types.

Finally, the algorithm should not be instrument-specific as this will severely restrict wide-scale application. To help achieve this, *treeseg* solely considers the 3D attributes of the plot-level point cloud. Inclusion of other attributes, for instance apparent reflectance, would introduce a requirement for some form of inter-instrument radiometric calibration. However, inevitably there will be some instrument-specific influence over the algorithm. The data considered in this thesis has been collected from a single high-end instrument whose prime concern is accurate and precise ranging. For low-end instruments, or for instruments where the ranging accuracy is not the priority, this mismatch will probably lead to poor method transferability.

5.2 Accessing the information content of point clouds

Prior to describing the methods of the *treeseg* algorithm in section 5.3, it is necessary to detail some of the underpinning generic point cloud manipulation techniques that permit access to the information content of a point cloud. To help illustrate these techniques, they have been performed on a section of plot-level point cloud acquired in plot NOU-11, as presented in figure 5.2. This point cloud, of dimensions 20 x 20 x 44 m LxWxH consists of 282 902 752 pts. As described in chapter 4 section 4.2.5, the plot-level point cloud of NOU-11 was obtained from the co-registration of scans from 121 locations using an angular resolution of 0.04°. Finally, noise reduction via deviation filtering, as described in chapter 4 section 4.1.6, has been performed on the plot-level point cloud.

Figure 5.2: Section of plot-level point cloud acquired in the tropical forests of Nouragues Nature Reserve, French Guiana (NOU-11), that is used in this section to demonstrate a number techniques that permit access to the information content of point clouds; coloured by height.

5.2.1 Nearest neighbour distance

The Euclidean distance, d, between the two points, p_1 and p_2, denoted in the usual Cartesian coordinate system, is defined as:

$$d(p_1, p_2) = \sqrt{(x_{p_1} - x_{p_2})^2 + (y_{p_1} - y_{p_2})^2 + (z_{p_1} - z_{p_2})^2} \tag{5.1}$$

The distance between the point p_1 and it's nearest neighbour, $d_{NN}(p_1)$, in the point cloud, P, $p \in P$, is defined as:

$$d_{NN}(p_1) = \min_{\{p \in P \,:\, p \neq p_1\}} d(p_1, p) \tag{5.2}$$

Across P, the mean nearest neighbour distance, $\overline{d_{NN}}$, is defined as:

$$\overline{d_{NN}} = \frac{1}{N(P)} \sum_{i=1}^{N(P)} d_{NN}(p_i) \tag{5.3}$$

In an idealised scene, such as the inside of a sphere with the scanner positioned at the centre, where the range from instrument to target, R, will be invariant, then for any arbitrarily selected point p_i from $p \in P$, $d(p_i) = \overline{d_{NN}}$. Therefore, $\overline{d_{NN}}$ describes the distance in P between topologically-connected points. That is, the average distance between two nearest points residing on the same surface. In this idealised scenario, $\overline{d_{NN}}$ will be directly related to the step-distance between sequentially fired pulses, which is defined by the angular resolution and R, as defined in chapter 3 section 3.1.5. The minimum of $\overline{d_{NN}}$ occurs at the smallest resolvable distance in the point cloud, d_{min}. As described in chapter 3 section 3.1, d_{min} is dictated by the fixed instrument-specific characteristics of ranging accuracy and beam spread.

As the complexity of the scene increases, such as for the sample data considered here, $\overline{d_{NN}}$ is influenced by two factors. First, variation in R across $p \in P$ will result in a variable step-distance and d_{min}. Secondly, the general distribution of elements in the forest will tend to increase occlusion more generally (e.g., significant quantities of vegetation in the undergrowth and understory will occlude the canopy elements). If it is assumed that the scene is sufficiently well populated, then provided $\overline{d_{NN}}$ is averaged over a sufficiently large volume to curtail any localised effects, it will describe the average distance in P between topologically-connected points as a function of these two influences.

In a forest scene, an increase in height from the forest floor goes hand in hand with an increase in the step-distance and d_{min} (i.e., increased R), and the impact of occlusion (i.e., all viewpoints are from close to the forest floor). Hence, the height-resolved average nearest neighbour distance, $\overline{d_{NN}}(z)$, will describe the average distance between topologically-connected points through the transition from forest floor to canopy. $\overline{d_{NN}}(z)$ is defined here using a bin at height, z, and width, dz, as:

$$\overline{d_{NN}}(z) = \frac{1}{N(P)} \sum_{i=1}^{N(P)} d_{NN}(p_i) \Bigg| \{p_i : z < z_{p_i} < z + dz\} \tag{5.4}$$

Figure 5.3 presents the distribution of $\overline{d_{NN}}(z)$ across the sample point cloud considered here using 1 m bins. It can be seen that over the first 0–30 m of height from the forest floor, the value of $\overline{d_{NN}}(z)$ varies smoothly over an order of magnitude. Above a height of about 30 m, $\overline{d_{NN}}(z)$ is more variable, which indicates that the influence of occlusion is becoming more prevalent. $\overline{d_{NN}}(z)$ is perhaps the most crucial parameter of the plot-level point cloud, in terms of its impact on the accuracy and reliability of *treeseg*. This is because the parameter is extensively used to drive a number of techniques described in this section, such as clustering algorithms that require an *a priori* description of the distance between topologically-connected points.

In addition to this, $\overline{d_{NN}}(z)$ can be used as a proxy for data quality. Data quality was described in chapter 3 section 3.1.5 as the completeness with which the plot-level point cloud captures the 3D structure of the scene, within the bounds of fixed instrument-specific parameters. In the canopy, if d_{min} falls below $\overline{d_{NN}}(z)$, then there is almost certainly scope to increase the sampling regime (e.g., scan pattern, scan number or angular resolution) to obtain a higher quality plot-level point cloud.

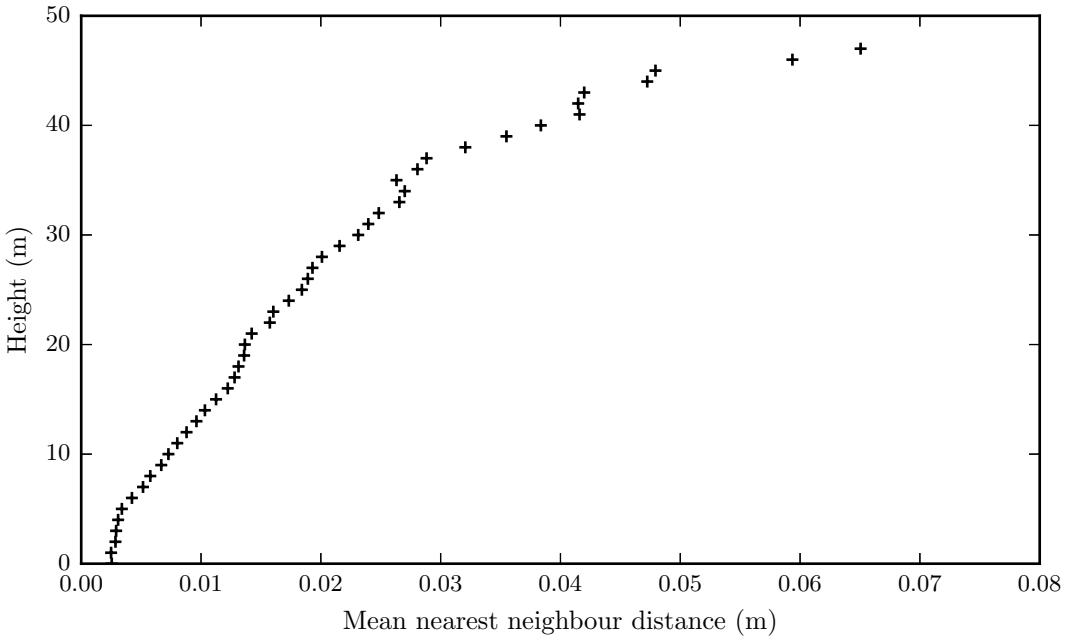


Figure 5.3: Height-resolved mean nearest neighbour distance, $\overline{d_{NN}}(z)$, across the sample point cloud comprising data from 121 scan locations.

5.2.2 Euclidean clustering

Spatial (Euclidean) clustering is a technique used to denominate a point cloud, P, based on point arrangement. This is undertaken by extracting a set of point clouds, $\{C\}$, from P, $\{C\} \subset P$. For each point cloud in $\{C\}$, the constituent points have a common nearest neighbour distance between at least one other point of less than or equal to d_{cmax}, and number more than or equal to N_{min}. A naive algorithm to generate the set $\{C\}$ is described in algorithm 1.

Algorithm 1: Naive Euclidean clustering.

Input: P, d_{cmax}, N_{min}

Output: $\{C\}$ ($\{C\} \subset P$)

Arbitrarily select point p_i from $p \in P$

    Determine $d(p_i, p)\{p \in P : p \neq p_i\}$

    Assign points with $d \leq d_{cmax}$ to C_1 and remove from P

    Repeat for all new $p \in C_1$

Place C_1 in $\{C\}$ if $N(C_1) \geq N_{min}$ else reinsert $p \in C_1$ into P

Repeat till exhaustion of $p \in P$

Figure 5.4 presents each cluster that has been derived through this algorithm from the sample point cloud when using a value of d_{cmax} slightly larger than $\overline{d_{NN}}$. Figure 5.5 then presents each cluster derived through this algorithm from the sample point cloud when using a value of d_{cmax} slightly larger than $\overline{d_{NN}}(z)$. These two figures demonstrate the influence the quantification of $\overline{d_{NN}}(z)$ can have on the performance of these clustering algorithms.

In the instance of figure 5.4, where $\overline{d_{NN}}$ is considered, it can be seen those clusters close to the forest floor (i.e., those closest to the scanner) are often large and contain many distinguishable surfaces. For instance, in the forefront, a large cluster coloured fluorescent green can be seen to contain points from the ground, stems and clumps of mixed branching and leafy vegetation. Whilst in the canopy, in comparison to the original data in figure 5.2, very few clusters have been identified. This is because the increasing d_{min} and influence of occlusion have not been compensated for.

Whereas, in figure 5.5 where $\overline{d_{NN}}(z)$ is considered, there is a more uniform distribution of clusters through the full extent of the point cloud. These clusters can be seen

to reasonably accurately represent their underlying surfaces, whether that be a single surface such as a stem, or a set of surfaces such as mix of branching and leafy material. This accurate extraction of clusters by their underlying surface is essential as they will be used as inputs for the feature extraction techniques that are described later in this section.

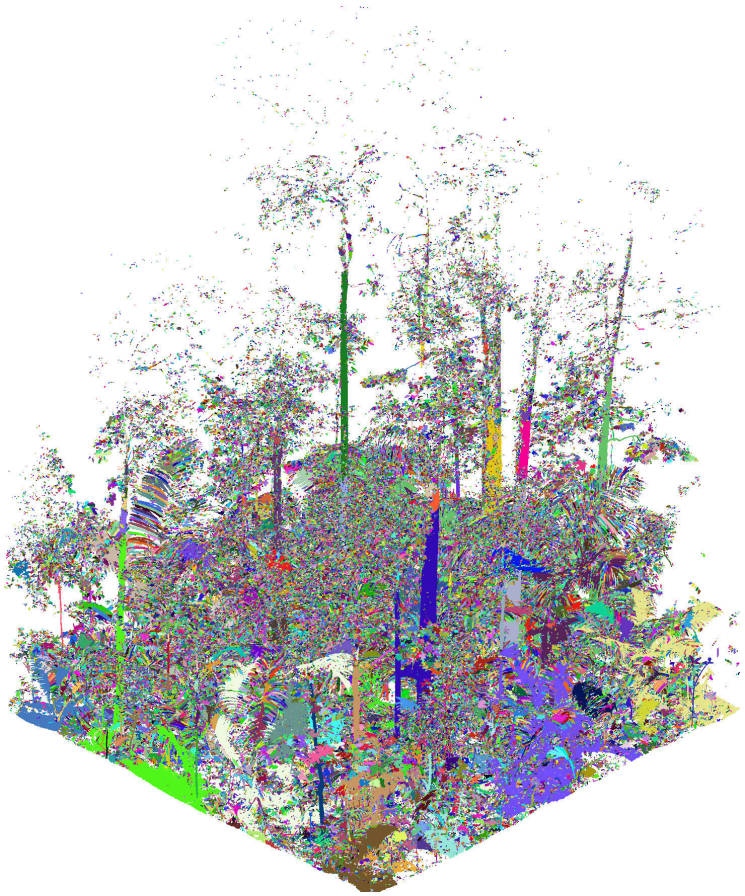


Figure 5.4: Uniquely coloured clusters extracted through Euclidean clustering from the sample point cloud when using a value of d_{cmax} slightly larger than $\overline{d_{NN}}$.

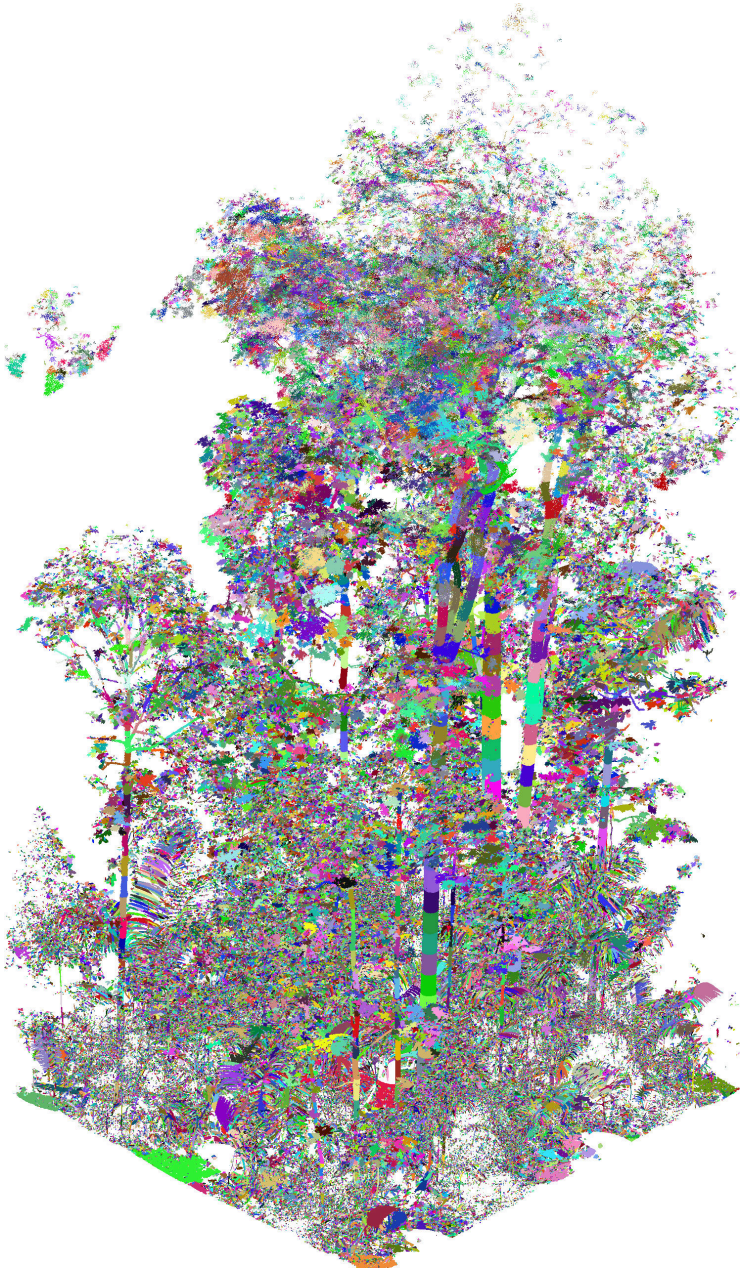


Figure 5.5: Uniquely coloured clusters extracted through Euclidean clustering from the sample point cloud when using a value of d_{cmax} slightly larger than $\overline{d_{NN}}(z)$.

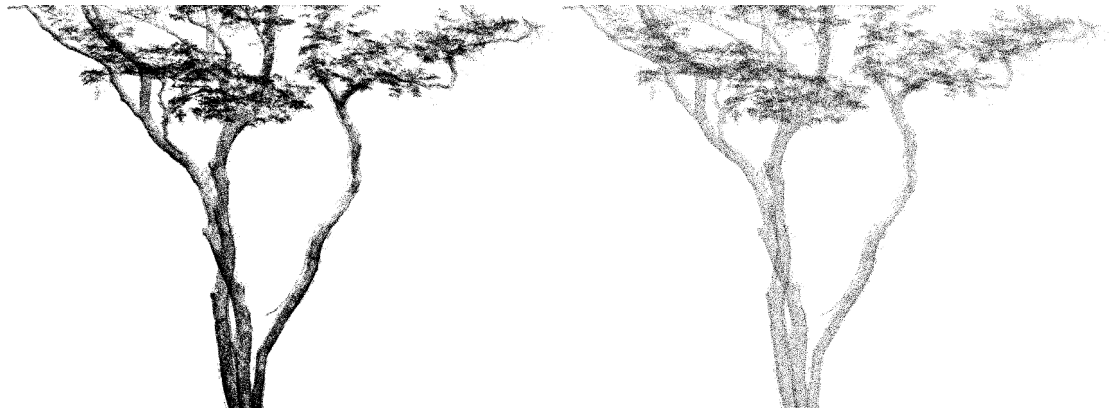
5.2.3 Downsampling

It was demonstrated in figure 5.5 that Euclidean clustering algorithms driven by $\overline{d_{NN}}(z)$ can be applied to accurately extract clusters representing topologically-connected surfaces. However, a significant limitation to this method is the inability to resolve localised variation in nearest neighbour distance (i.e., intra-bin). This localised variation could arise in significantly occluded areas such as around liana infestations, or be a result of an asymmetric sampling regime.

A robust method to reduce any localised variation is point cloud downsampling. Downsampling can be achieved via voxelisation. That is, using cubes of constant edge length, l_V, to partition the full extent of the point cloud. Points lying inside each voxel, $p \in V$, are then aggregated into their centroid point, p_c, defined as:

$$x_{p_c} = \frac{1}{N(V)} \sum_{i=1}^{N(V)} x_{p_i} \qquad y_{p_c} = \frac{1}{N(V)} \sum_{i=1}^{N(V)} y_{p_i} \qquad z_{p_c} = \frac{1}{N(V)} \sum_{i=1}^{N(V)} z_{p_i} \tag{5.5}$$

From the sample point cloud, figure 5.6a shows a sub-section of this cloud that represents a section of canopy branching. This extracted point cloud has then been downsampled, as shown in figure 5.6b, through voxel grid downsampling where $l_V = 0.04\,\text{m}$. It can be seen that as a result of the downsampling, the distribution of points across the point cloud has become significantly more uniform with any localised concentrations of points being dispersed. This leads to the visual appearance of the point cloud being much less 'patchy'.



(a) **(b)**

Figure 5.6: (a) Section of point cloud representing some canopy branching that has been extracted from the sample point cloud, and (b) the section of point cloud after being downsampled via voxel grid downsampling where $l_V = 0.04\,\text{m}$.

Returning to the sample point cloud, figure 5.7 presents the distribution of $\overline{d_{NN}}(z)$ across the sample point cloud before and after voxel grid downsampling where $l_V = 0.04\,\text{m}$. Across the first 30 m of height through the plot, it can be seen that the gradient of $\overline{d_{NN}}(z)$ in the downsampled point cloud is significantly steeper in comparison to the original point cloud. Through this first 30 m of height, the full range of $\overline{d_{NN}}(z)$ has also decreased from 0.002–0.023 m to 0.022–0.034 m, representing a 42.9 % reduction.

This explicitly demonstrates that the variation in d_{NN} through P from the forest

floor to the canopy has been reduced because of the downsampling. Implicit in this must be the reduction in localised variation of d_{NN}. It should be noted in figure 5.7 that the minimum $\overline{d_{NN}}(z)$ does not equal l_V. This is because the arrangement of $p \in V$ are structured by underlying surfaces and not randomly distributed. Thus, the distance between aggregated (centroid) points will be both variable and more closely aligned to the underlying surfaces.

The benefit of downsampling in creating a more uniform $\overline{d_{NN}}(z)$ can be seen in figure 5.8, where the aforementioned Euclidean clustering algorithm has been applied to the downsampled point cloud using a value of d_{cmax} slightly larger than $\overline{d_{NN}}(z)$. In comparison to figure 5.5, it can be seen that the derived clusters are more uniform and less 'patchy'. That is, they more distinctly partition the underlying surfaces whilst loss of points in the upper canopy is reduced noticeably.

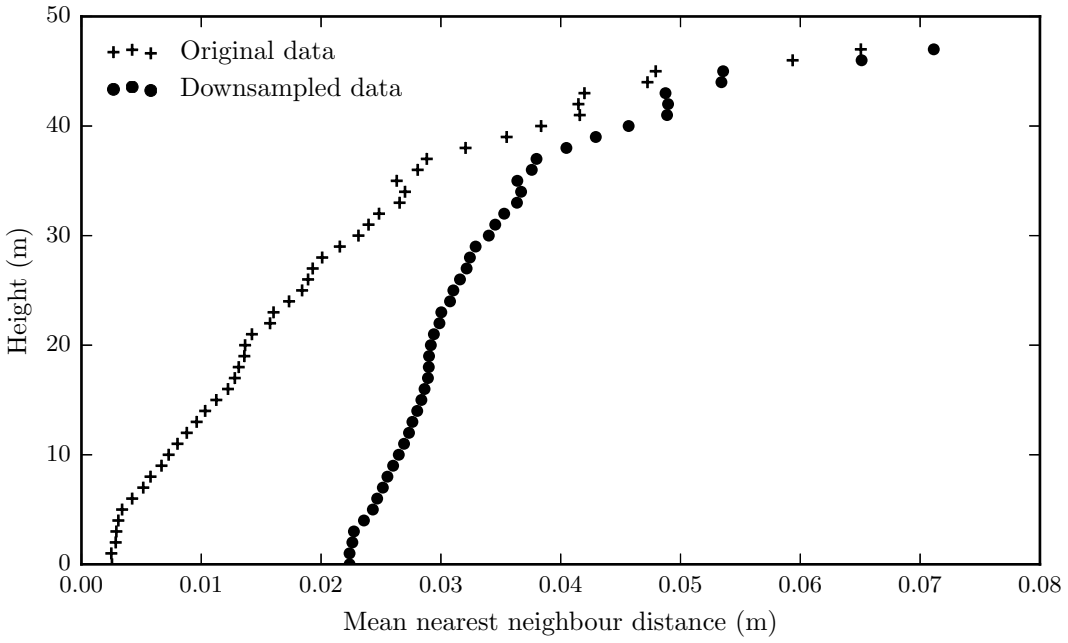


Figure 5.7: Height-resolved mean nearest neighbour distance, $\overline{d_{NN}}(z)$, across the sample point cloud and subsequently downsampled sample point cloud using voxel grid downsampling where $l_v = 0.04\,\mathrm{m}$.

Downsampling then, plays a significant role in reducing the magnitude of variation in $\overline{d_{NN}}(z)$ through a point cloud. It also plays a significant role in subduing localised variation in nearest neighbour distance. As the algorithms considered in this section, such as Euclidean clustering are shown to be sensitive to these influences, downsampling is a key requirement for their consistent performance.

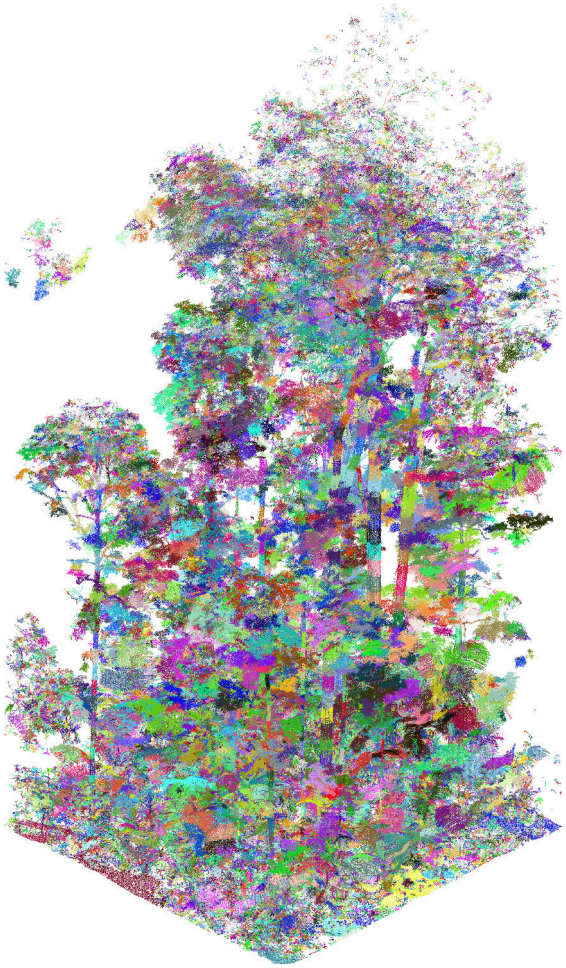


Figure 5.8: Uniquely coloured clusters extracted through Euclidean clustering from the downsampled sample point cloud when using a value of d_{cmax} slightly larger than $\overline{d_{NN}}(z)$, the downsampling has been undertaken through voxel grid downsampling where $l_V = 0.04\,\text{m}$.

However, downsampling can of course play a potentially negative role in retrieving information from a point cloud. The definition of l_V is, all other things being equal, a redefinition of d_{min}. That is, if the value of l_V is large enough, then the resolution of the point cloud will be reduced. This could, amongst other impacts, lead to apparent increases in the size of woody structure that would induce a systematic bias in products derived further down the line, such as tree-scale volume. Selection of l_V must be a trade-off between establishing a more uniform $\overline{d_{NN}}(z)$ whilst having minimal impact on reducing the resolution of the point cloud.

Chapter 3 section 3.1 provided a definition of the diameter of the footprint at R. In the context of the sample point cloud considered here, through the first 10 m of height, it might not be unusual for an average value of R to be around 15 m. Using the instrument parameters defined in chapter 4 section 4.1.1, this would lead to a diameter of an average

beam footprint being 0.012 m. Thus, a value of $\overline{d_{NN}}(z)$ below this value is essentially indicative of oversampling, as the distance between the points is ambiguous. As can be seen in figure 5.7, when $l_v = 0.04\,\text{m}$, a sizeable portion of the points pruned during the downsampling come from this oversampled first 10 m.

Finally, a second but relevant consequence of downsampling is compute time. The downsampled point cloud has a point count of 17 257 067 pts, which represents a 94 % reduction over the original data. To the naive routines described so far, this will correspondingly reduce the compute time by $\approx 90\,\%$.

5.2.4 Principal component analysis

In the context of 3D point clouds, principal component analysis (PCA) is a technique that can be used to summarise the arrangement of points in some point cloud, P (Pauly et al. 2003). PCA permits P to be reduced to three vectors that describe the direction and magnitude of correlation between x, y and z for $p \in P$. This is achieved by Eigen decomposition, defined as:

$$(C - \lambda I)v = 0 \tag{5.6}$$

Where, λ and v are the Eigenvectors and Eigenvalues respectively. C, the covariance matrix, is defined as:

$$C = \begin{bmatrix} \text{cov}(x,x) & \text{cov}(x,y) & \text{cov}(x,z) \\ \text{cov}(y,x) & \text{cov}(y,y) & \text{cov}(y,z) \\ \text{cov}(z,x) & \text{cov}(z,y) & \text{cov}(z,z) \end{bmatrix} \tag{5.7}$$

Where, $\text{cov}(x,y)$ for instance, is defined as:

$$\text{cov}(x,y) = \sum_{i=1}^{N(P)} \frac{(x_{p_i} - x_{p_c})(y_{p_i} - y_{p_c})}{N(P) - 1} \tag{5.8}$$

Definition of C in equation 5.6, presents a cubic polynomial that can be solved for λ by factorisation. Subsequently, v can be solved from each respective resultant linear equation.

To illustrate the purposes of PCA here, the point cloud shown in figure 5.9 represents a section of stem that has been extracted from the sample point cloud. This point cloud

has then been downsampled via voxel grid downsampling where $l_V = 0.04\,\text{m}$. At the centre of the point cloud, the primary Eigenvector λ_1, secondary Eigenvector, λ_2 and tertiary Eigenvector, λ_3, have been overlaid, scaled by the relative magnitude of their respective Eigenvalues, v_1, v_2 and v_3.

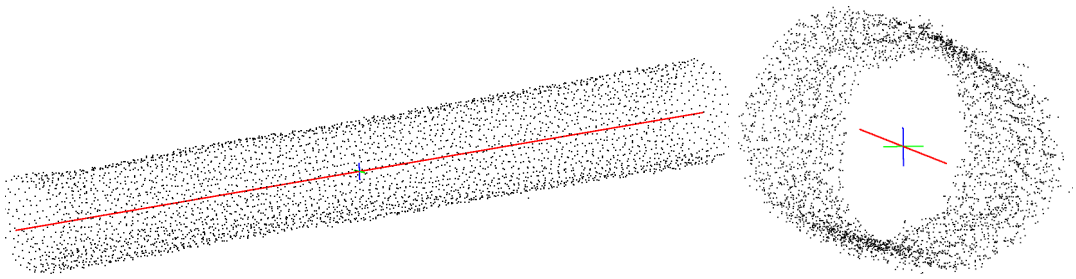


Figure 5.9: Section of stem ($D = 0.67\,\text{m}$), extracted and downsampled from the sample point cloud; PCA has been performed on this section and the three derived Eigenvectors, scaled by their respective Eigenvalues, have been overlaid (red, blue, green); view of same object from different perspectives and at different scales.

It can be seen that the Eigenvalue of the primary Eigenvector is sufficiently larger than that of the other two Eigenvectors, that it can be said the primary Eigenvector describes the majority of the correlation between x, y and z for $p \in P$. As a result, the primary Eigenvector provides a reasonable interpretation of the directionality of the stem. This interpretation can be powerful when applied across multiple sets of points. This is because it offers derived heuristics, such as the angle between the primary Eigenvectors, to infer, amongst other things, connectivity.

This is illustrated in figure 5.10, where a point cloud representing a section of branching has been extracted from the sample point cloud. After downsampling via voxel grid downsampling where $l_V = 0.04\,\text{m}$, Euclidean clustering has then been applied to this point cloud using a value of d_{cmax} slightly larger than $\overline{d_{NN}}(z)$. Subsequently, PCA has been performed on each resultant cluster with the derived principal Eigenvector overlaid on each cluster.

It can be seen that over the woody clusters, the principal Eigenvector produces a reasonable skeletal outline of the branching. One application of this skeleton could be to use the distance and angle between Eigenvectors to determine if the clusters are connected, and resultantly belong to the same tree.

However, towards the high order branching, where the clusters comprise a mixture of returns from wood and leaf targets, this skeletal definition fades. This is a direct result of the arrangement of the points in these clusters. Significantly less structured

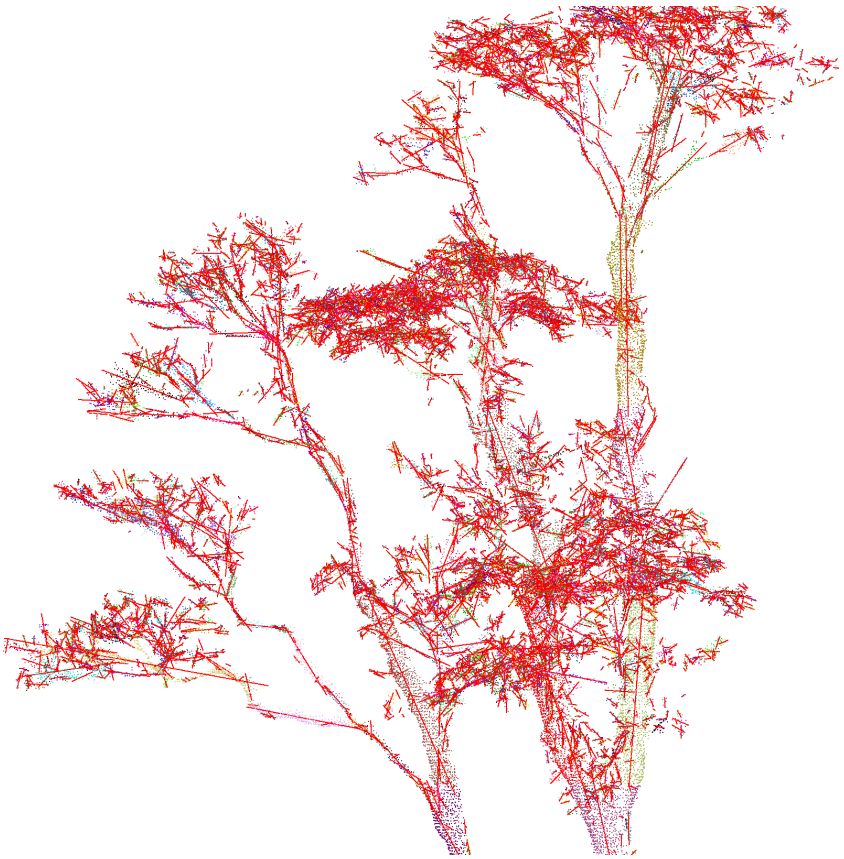


Figure 5.10: Section of the sample point cloud, extracted and downsampled from the sample point cloud, Euclidean clustering has then been applied to the point cloud with each derived cluster uniquely coloured; PCA has been performed on each cluster with the derived λ_1 overlaid on each cluster in red.

than their branch cluster counterparts, they appear more 'fuzzy'. By definition, this increased randomness in their arrangement means there is less correlation between x, y and z for $p \in C$. This results in the magnitude between v_1, v_2 and v_3 reducing such that the appearance of λ_1 across these clusters is less coherent.

5.2.5 Surface normal estimation

One particular extension of PCA is surface normal estimation. Across a set of points, S, obtained from a common surface, such as a leaf, the normal to this surface is estimated by the definition of the vector perpendicular to the plane fitted through S. By definition, λ_1 across S, derived via PCA, is the vector of this plane. As each Eigenvector is orthogonal to the next, λ_3 defines the normal of this plane (i.e., an estimate of the surface normal).

Surface normals can be estimated for every point in a point cloud, such as an example point cloud, P. However, prior to surface normal estimation for the first arbitrarily selected point p_i from $p \in P$, it is necessary to define a search radius, r_N. Points

neighbouring p_i that fall inside this search radius constitute the set, S. PCA is then performed on S, with λ_3, the estimated surface normal, attributed to p_i.

To demonstrate surface normal estimation, a point cloud representing a section of branching has been extracted from the sample data and downsampled via voxel grid downsampling where $l_V = 0.04\,\text{m}$. Figures 5.11a and 5.11b present this point cloud, shaded by the orientation of the estimated surface normals, where $r_N = 2\overline{d_{NN}}(z)$ and $r_N = 6\overline{d_{NN}}(z)$ respectively. It should be noted that the sign of the estimated surface normal is always unknown, so in this instance all normals are oriented about the position from which the clouds are viewed.

(a) **(b)**

Figure 5.11: Section of tree structure, extracted and downsampled from the sample point cloud, where each point is shaded by the orientation of the estimated surface normal that has been generated across the neighbouring set, defined in (a) by $r_N = 2\overline{d_{NN}}(z)$ and (b) as $r_N = 9\overline{d_{NN}}(z)$.

It can be seen in figure 5.11b that the orientations of the estimated surface normals across the tree structure are fairly consistent. This consistency could be used to infer that all these points belong to the same underlying surface and as such, could potentially be used to crudely filter out this section of branching. In figure 5.11a, where a smaller search radius has been used, the orientation of the estimated surface normals appear more random. This is because the underlying surface, the branching, is cylindrical. Therefore the smaller sized S, whose underlying surface captures only a small proportion of the cylinder curvature, leads to non-consistent orientation of the surface normal vectors across the branching.

Finally, one further interesting characteristic of the underlying surface that can be estimated concurrent with the surface normals is curvature. That is, the goodness of the fitted plane through S can provide an estimate of the flatness of the underlying surface.

Curvature, c, is defined as:

$$c = \frac{\lambda_1}{\lambda_1 + \lambda_2 + \lambda_3} \tag{5.9}$$

Such that when λ_1 dominates over λ_2 and λ_3, the plane is well-defined and the underlying surface could be estimated as flat. As the relative magnitude of λ_2 and λ_3 increase, the underlying surface becomes non-flat (i.e., somehow curved).

Figure 5.12 presents the same section of tree structure as figure 5.11 that has been coloured by the value of curvature defined for each point by the neighbouring set determined by $r_N = 9\overline{d_{NN}}(z)$. Along the most pronounced sections of branching, there is some consistency in the value of curvature, again indicating that these points may share the same common surface. However around the high order branching and leafy material this consistency stops. As described in section 5.2.4, this is a result of S containing a mixture of points from wood and leafy clumps. This random arrangement will lead to inconsistent magnitudes in the Eigenvalues, causing a much wider, less systematic distribution of curvature.

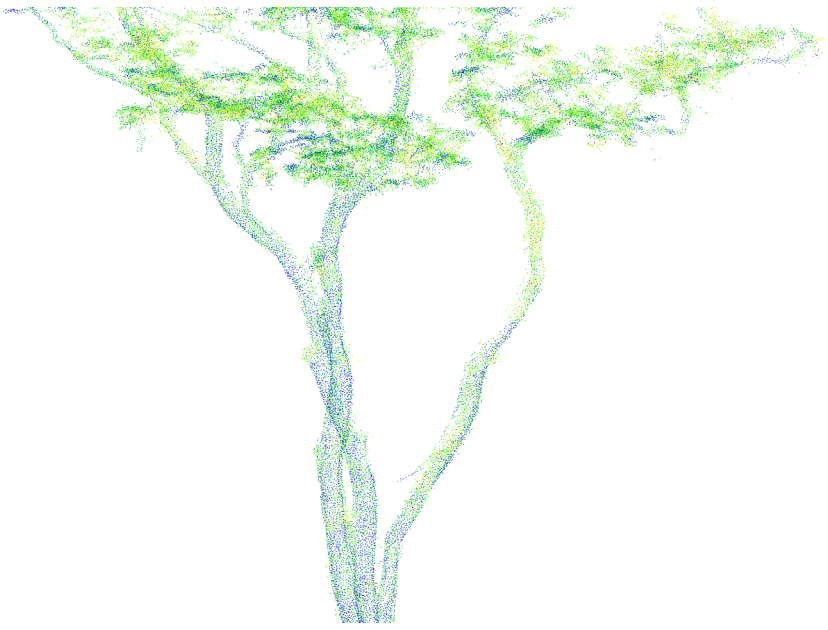


Figure 5.12: Section of branching, extracted and downsampled from the sample point cloud, where each point is coloured by the value of curvature that has been generated across the neighbouring set defined by $r_N = 9\overline{d_{NN}}(z)$.

5.2.6 Region growing segmentation

An application of PCA and surface normal estimation is region growing segmentation (Rabbani et al. 2006). The objective of region growing segmentation is to partition a point cloud, P, into a set of regional point clouds, $\{R\}$, $\{R\} \subset P$. The constituent points in each of these regional point clouds, $R \in \{R\}$, will have some degree of commonality such that an inference can be made that they belong to the same underlying surface.

To determine if two points share a common surface, two metrics are used in region growing segmentation. First, the aforementioned value of curvature described in the previous section, and second, the angle between the surface normals. The angle between two vectors, A and B, is defined as:

$$\theta(A, B) = \arccos\left(\frac{A \cdot B}{|A||B|}\right) \tag{5.10}$$

A decision on commonality is made using the hard thresholds, c_{max} and θ_{max}, that define the maximum tolerable divergence in curvature and angle between surface normals respectively. A naive region growing algorithm to generate the set $\{R\}$ is described in algorithm 2.

Algorithm 2: Naive region growing.

Input: P, r_N, c_{max}, θ_{max}

Output: $\{R\}$,($\{R\} \subset P$)

1 For $p \in P$ determine N_p and c_p using the neighbouring set defined by r_N

2 The seed point p_s is identified by $\min c_p$ for $p \in P$

3 Search for points surrounding p_s and insert into point cloud P_{tmp}

4 For $p \in P_{tmp}$ if $\theta(N_s, N_p) \leq \theta_{max}$ and $|c_p - c_s| \leq c_{max}$ add to region cloud R_1

5 Repeat for all new $p \in R_1$

6 Remove $p \in R_1$ from P and place R_1 in $\{R\}$

7 Repeat till exhaustion of $p \in P$

To demonstrate this algorithm, a section of tree structure has been extracted from the sample point cloud and downsampled via voxel grid downsampling where $l_V = 0.04\,\text{m}$. Region growing segmentation has then been performed on the section where $r_N = 9\overline{d_{NN}}(z)$, $c_{max} = 0.75$ and $\theta_{max} = 20°$. Figure 5.13 presents each of the region

point clouds that partition the section of tree structure with a unique colour. It can be seen that the most prominent section of branching belongs to a single surface. Just as importantly, the higher order branching and the clumps of leafy material have been segmented into their own distinct regional point clouds. That is, there are few point clouds in $\{R\}$ that represent two separate underlying surfaces. This establishes that region growing segmentation can partition the plot-level point cloud into point clouds belonging to large sections of woody structure, small sections of woody structure and clumps of leafy vegetation.

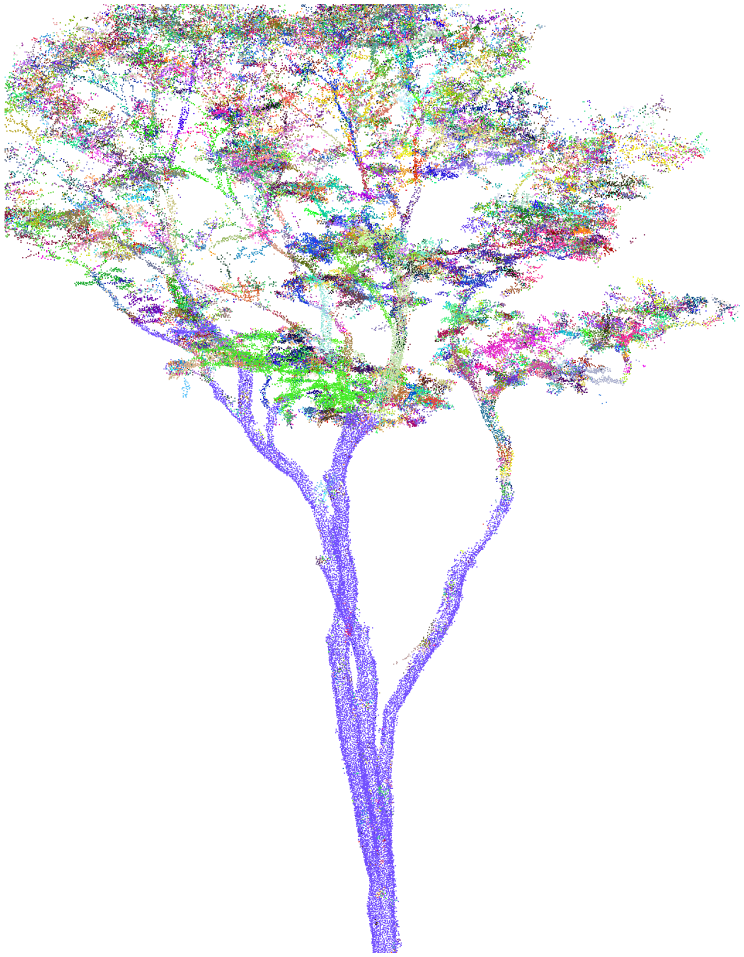


Figure 5.13: Section of tree structure, extracted and downsampled from the sample point cloud, where each uniquely coloured regional point cloud have been segmented using region growing segmentation where $r_N = 9\overline{d_{NN}}(z)$, $c_{max} = 0.75$ and $\theta_{max} = 20°$

However, the caveat to region growing segmentation is the requirement for an *a priori* definition of the hard thresholds c_{max} and θ_{max}. The values used in the above example have been derived through trial and error. Similar to the sensitivity of Euclidean clustering to variation in nearest neighbour distance, region growing segmentation is likewise sensitive to the choice of these parameters. As discussed in section 5.2.1, this

will inevitably be present in TLS data acquired from forest plots. As a result, the use of downsampling is essential to alleviate this sensitivity to variation in nearest neighbour distances.

5.2.7 Random consensus model fitting

The final method described in this section is random consensus model fitting (RANSAC) (Schnabel et al. 2007). RANSAC methods are used to optimise the fit of a geometric primitive to a point cloud, P. The geometric primitive considered here is a cylinder.

A cylinder is defined by three parameters; first, a centreline with vector, v, second, a point on the centreline, p_0, and third, a radius, r. For a cylinder fitted to P, the vector of the cylinder centreline is identical to the vector of the plane fitted through P. That is, by employing PCA, v can be equated to λ_1. This plane can be defined by the equation:

$$\lambda_{(1,x)}x + \lambda_{(1,y)}y + \lambda_{(1,z)}z + -\lambda_{(1,x)}x_{p_c} - \lambda_{(1,y)}y_{p_c} - \lambda_{(1,z)}z_{p_c} = 0 \tag{5.11}$$

Where p_0 can be described is the centroid point, p_c, of P, as defined in equation 5.5.

Following this, r of the cylinder fitted through P is defined as the mean distance between $p \in P$ and the plane. This shortest distance between some arbitrarily selected point, p_1, and the plane, d, is defined as:

$$d = \frac{|\lambda_{(1,x)}x_{p_1} + \lambda_{(1,y)}y_{p_1} + \lambda_{(1,z)}z_{p_1} - \lambda_{(1,x)}x_{p_c} - \lambda_{(1,y)}y_{p_c} - \lambda_{(1,z)}z_{p_c}|}{\sqrt{\lambda^2_{(1,x)} + \lambda^2_{(1,y)} + \lambda^2_{(1,z)}}} \tag{5.12}$$

If the cylinder fitted to P is to be finite then it must be characterised by one further parameter, length, l. To determine l from P, it is necessary to transform λ_1 such that the centreline z-axis will be rotated into (0,0,1). This transformation can be achieved through the rotation:

$$\begin{bmatrix} x' \\ y' \\ z' \end{bmatrix} = \begin{bmatrix} \cos\theta & \sin\theta & 0 \\ -\sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} \tag{5.13}$$

Where θ, the angle between λ_1 and (0,0,1) is calculated through equation 5.10. Through

this rotation of P, l can be defined as:

$$l = p_{z_{max}} - p_{z_{min}} \tag{5.14}$$

Where $p_{z_{max}}$ and $p_{z_{min}}$ are the largest and smallest value of $p_z \in P$.

So these equations define the fitting of a cylinder model, with parameters v, p_0, r and l to the point cloud P. The RANSAC method is used to optimise this fit across P, where P may contain either noisy points or points laying outside of the cylindrical surface in question. The method is a simple brute-force iterative resampling of P with replacement to produce the N-sized set of point clouds $\{R\}$. For each point cloud in the set, $R \in \{R\}$, the cylinder model is fitted to the constituent points to produce N cylinder models. The goodness of each of these cylinder models can be determined by the residual error of the radius, σ_r, applied across P, defined as:

$$\sigma_r = \sqrt{\frac{\sum_{i=1}^{N(P)} (d(p_i) - r)^2}{N(P)}} \tag{5.15}$$

The model with the minimum value of σ_r is considered the optimised cylinder model. The inliers, the points belonging to the point cloud R from which the optimised cylinder model was generated are extracted from P and into an inlier point cloud.

This method is demonstrated in figure 5.14, where a section of point cloud has been extracted from the sample point cloud. This section of point cloud contains a portion of stem and a large palm leaf and has been downsampled via voxel grid downsampling where $l_V = 0.04\,\text{m}$. The figure illustrates the inliers (red) and outliers (black) of the optimised cylinder model, derived via a RANSAC cylinder fit where $N = 1000$.

It can be seen that the optimised cylinder fit has correctly identified the inlier points belonging to the stem. As a result, the palm leaf returns have been segmented out. This demonstrates the ability of RANSAC methods to prune outliers even when they are of significant number and belong to a separate underlying structured surfaces (i.e., not just random noise).

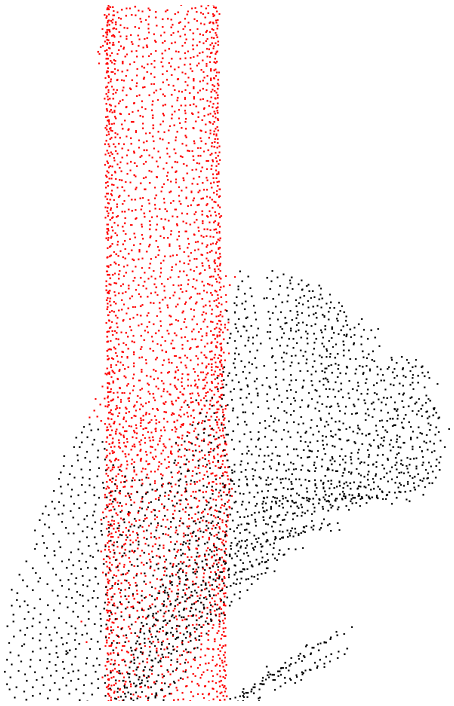


Figure 5.14: A section of point cloud extracted and downsampled from the sample point cloud that contains a section of stem and a palm leaf; a RANSAC cylinder fit has been performed on the point cloud where $N = 1000$, the model inliers and outliers have been coloured red and black respectively.

5.3 Methods

This section describes the methods of the *treeseg* algorithm. The objective is to ingest a plot-level point cloud, P, and near-automatically extract a set of tree-level points clouds, $\{T\}$, $\{T\} \subset P$. This is achieved by assessing the structure of the underlying surface of intentionally partitioned sections of P through Euclidean clustering, region growing segmentation, PCA and RANSAC cylinder model fitting. A high-level overview of *treeseg* is provided in algorithm 3:

Algorithm 3: Overview of *treeseg*.

Input: P

Output: $\{T\}$ ($\{T\} \subset P$)

1 Downsample P to produce the downsampled plot-level point cloud, P_d

2 Generate a digital elevation model across P_d

3 Create a slice, P_s, through P_d in the z-axis guided by the digital elevation model

4 Extract a set of point clouds, $\{R\}$, from P_s that have been identified as stems

For each identified stem $R \in \{R\}$:

5 Extract the stem point cloud, S, from P_d up to the position of first branching

6 Extract the crown point cloud, T, from the local section of canopy in P_d

7 Where necessary, manually refine T

8 Retrieve the tree-level structural parameters of x, y, D, H and C_r from T

The methods for each of the steps 1-8 in algorithm 3 are described in detail in the following sections. To demonstrate these methods, they have been applied to the plot-level point cloud acquired in plot NOU-11, Nouragues Nature Reserve, French Guiana. As described in chapter 4 section 4.2.5, the forest in NOU-11 represented the most structurally complex encountered during the collection of TLS data for this thesis. This was largely a result of the high stem count, the significant interaction between vegetation across the forest layers and considerable number of *Arecaceae* family plants densely packed through the undergrowth and understory.

To apply these methods, the Point Cloud Library (PCL) (Rusu et al. 2011), has been used. Implemented in C++, this open source library offers numerous generic algorithms that facilitate access to the information content of points clouds in a fast and efficient

manner.

5.3.1 Data pre-processing

treeseg has three expectations of P. First, the data that drove the development of the algorithm has been acquired from an instrument with a ranging accuracy of 5 mm over several 100 m. It is unlikely that the algorithm will be transferable in its entirety to data acquired by instruments with significantly lower ranging accuracy.

Second, the accuracy in the co-registration of scan-level point clouds into P should be of the same order as the ranging accuracy of the instrument. As demonstrated in chapter 4 section 4.1.4, for the data considered here, the two-step co-registration process permits the scan-level point clouds constituting P to be indistinguishable from one another.

Finally, noise in P should be removed where possible. All points inside P are treated equally, so the presence of significant noise can be expected to interfere with the extraction. As illustrated in chapter 4 section 4.1.6, for the data considered here, noise reduction was undertaken via deviation filtering.

5.3.2 Downsampling

Step one of the *treeseg* algorithm is to downsample P into the downsampled plot-level point cloud, P_d. This is undertaken through voxel grid downsampling, as described in section 5.2.3. This is to to provide a more uniform distribution of $\overline{d_{NN}}(z)$ from the forest floor through to the canopy, as well as to subdue localised variation in nearest neighbour distance throughout P.

For all the data considered here, $l_V = 0.04$ m. Figure 5.15 presents the distribution of $\overline{d_{NN}}(z)$ across both P and P_d, alongside the diameter of the beam footprint as a function of height. As a result of this downsampling the point count in P reduced from 4 269 712 380 pts to 338 444 016 pts, representing a 92 % reduction.

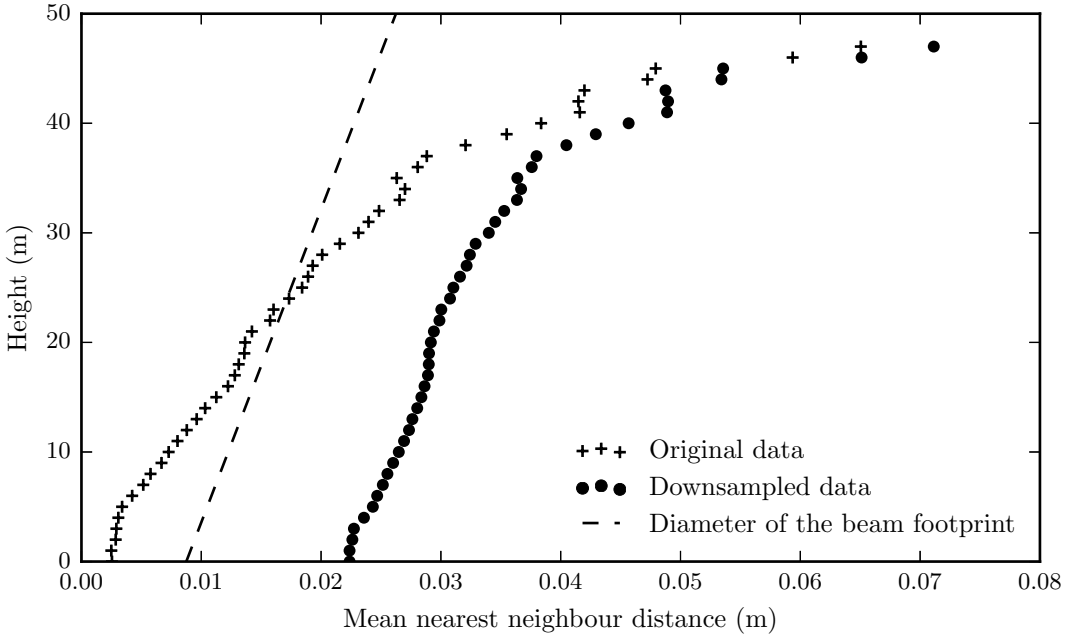


Figure 5.15: Height-resolved mean nearest neighbour distance, $\overline{d_{NN}}(z)$, across P and P_d obtained via voxel grid downsampling where $l_v = 0.04\,\text{m}$; the diameter of the beam footprint as a function of height is also shown.

5.3.3 Digital elevation model

Post-downsampling, the second step of *treeseg* is to generate a digital elevation model (DEM) of the plot. This is achieved by partitioning the 2D bounding box (x-y) of P_d into voxels of edge length, r. The minimum point inside each voxel represents the forest floor, z_{floor}, defined as:

$$z_{floor} = \min_{\{p \in V\}} z_p \Big| \{p_i : x < x_{p_i} < x + r\}\{y_i : y < y_{p_i} < y + r\} \tag{5.16}$$

The DEM is then formed by generating an interpolated mesh across z_{floor} for each voxel. Figure 5.16 presents the DEM derived from P_d where $r = 2\,\text{m}$. This coarse resolution is acceptable as only an approximate DEM is required by the methods in the following sections. A low resolution also negates the possibility of a large stem entirely occupying a voxel that would lead to a false estimation of the position of the forest floor.

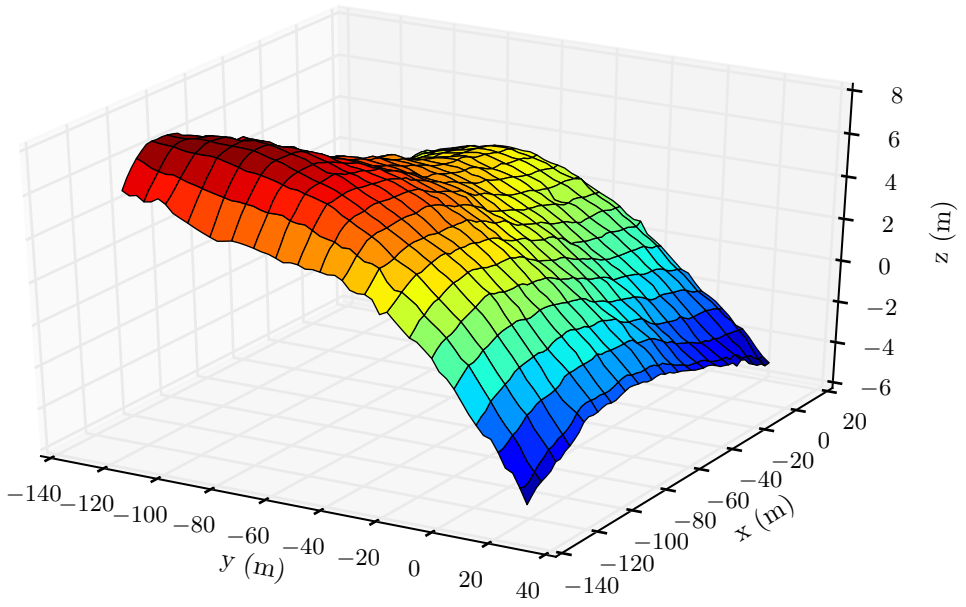


Figure 5.16: DEM derived across P where $r = 2\,\text{m}$.

5.3.4 Stem identification

The third and fourth steps of *treeseg* are to identify the location of the stems across P_d. This is achieved by creating a slice, P_s, across P_d in the z-axis, between 3–6 m. The DEM is used to guide the position at which the slice is taken through P_d to compensate for the topography of the forest floor. P_s is presented in figure 5.17.

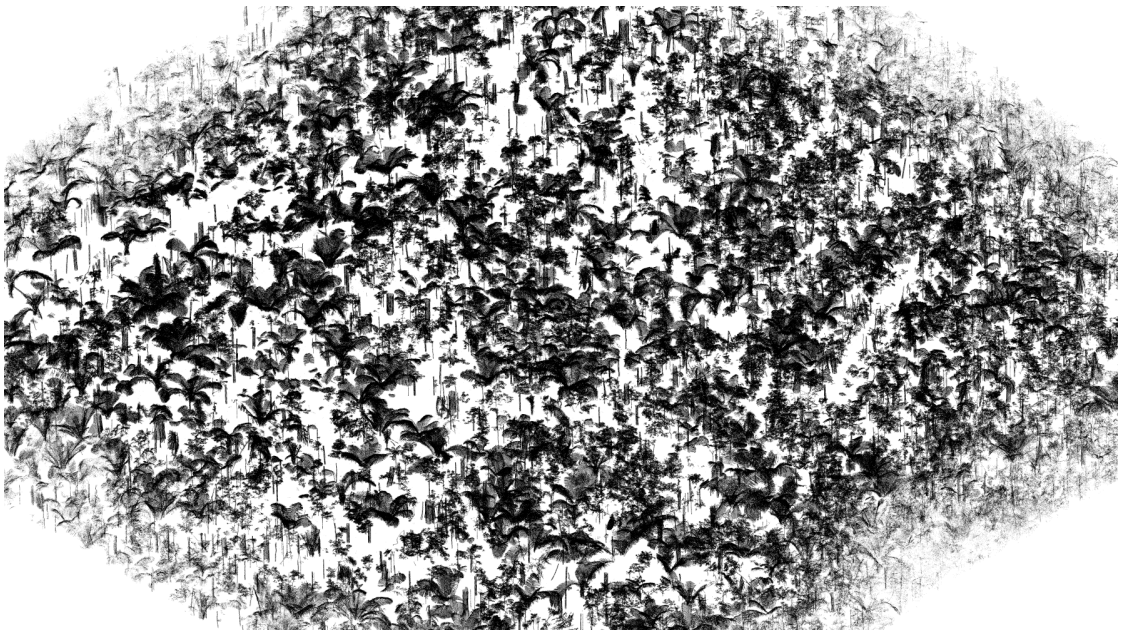


Figure 5.17: Slice, P_s, through P_d in the z-axis between 3–6 m as guided by the DEM.

Following this, P_s is partitioned into a set of point clouds, $\{C\}$, $\{C\} \subset P_s$, each representing topologically-connected surfaces. For each point cloud in this set, $C \in \{C\}$, the topologically connected surfaces may be a single surface such as a stem or a mix of inter-connected surfaces such as clumps of vegetation and branching. This partitioning is undertaken through Euclidean Clustering, as described in section 5.2.2. Here, $d_{cmax} = 9\overline{d_{NN}}(z)$ and $N_{min} = 10$. Figure 5.18 presents the set $\{C\}$ with each constituent point cloud uniquely coloured.

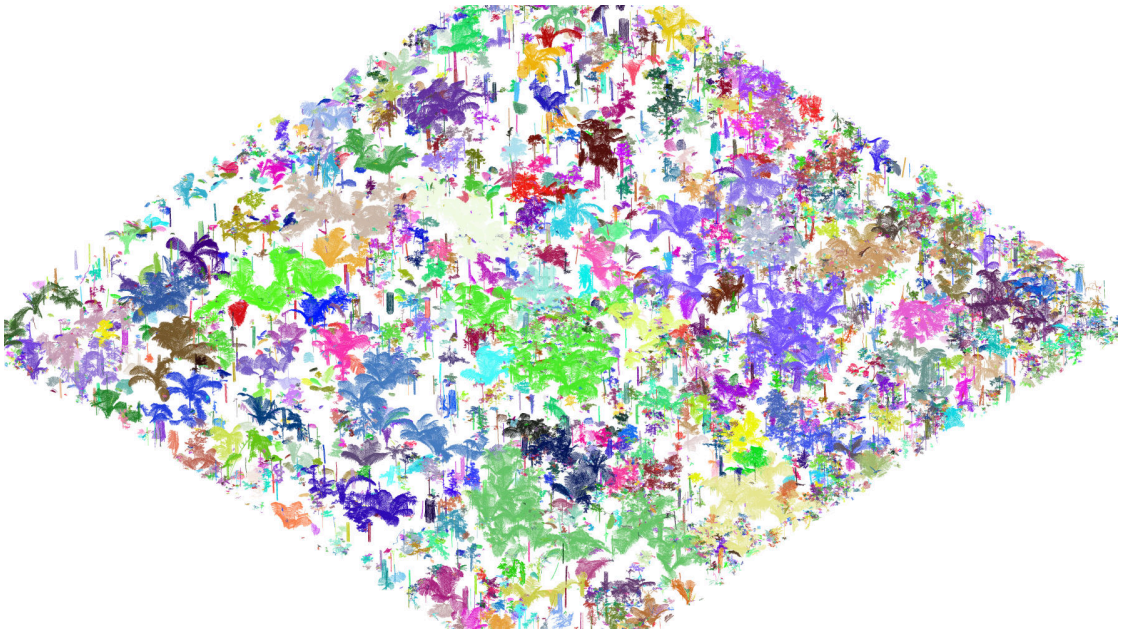


Figure 5.18: The point clouds, each uniquely coloured, that comprise the set $\{C\}$, partitioning P_s via Euclidean Clustering to represent topologically-connected surfaces.

Next, each point cloud comprising $\{C\}$ is further partitioned into the set of point clouds, $\{R\}$, $\{R\} \subset \{C\}$,that each solely represent topologically-connected single surfaces. This further partitioning is undertaken through region growing segmentation, as described in section 5.2.6. Here, $r_N = 9\overline{d_{NN}}(z)$, $c_{max} = 1$ and $\theta_{max} = 25°$. Figure 5.19 presents the set $\{R\}$ with each constituent point cloud uniquely coloured.

The next step is to determine whether the point clouds in the set $\{R\}$ represent either a stem or some other element of vegetation. This determination is undertaken through RANSAC cylinder fitting, as described in section 5.2.7. For each point cloud in the set, $R \in \{R\}$, R is segmented into five equal subsections by filtering in the z-axis. For each of these sections and for R itself, a cylinder model is fitted using the RANSAC approach with $N = 1000$.

Two metrics are then used to assess whether R represents a stem. First, the coefficient

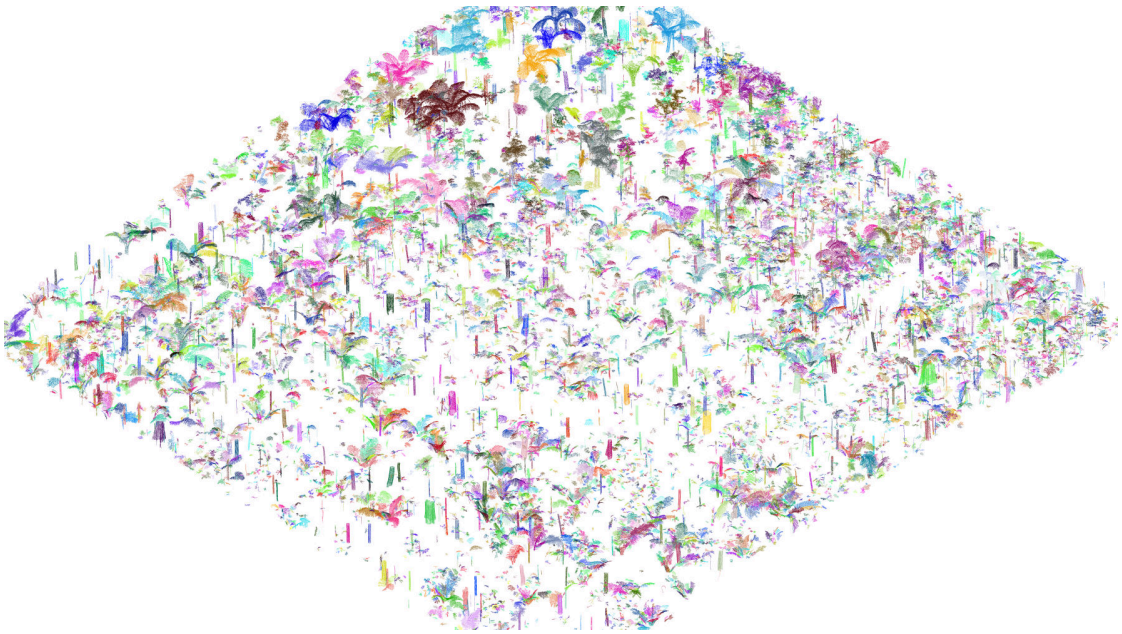


Figure 5.19: The point clouds, each uniquely coloured, that comprise the set $\{R\}$, further partitioning the set $\{C\}$ via region growing segmentation to represent topologically-connected single surfaces.

of variation, CV, across the radii derived from the five subsections. Second, the ratio between the radius of the cylinder model fitted to R and the mean radius from the five cylinder models fitted to the subsections, r_r. These two metrics permit questioning of the stability of the fitted cylinder model, and as such can be interpreted as how cylindrical the underlying surface is.

Here, the thresholds $CV_{max} = 0.2$ and $r_{r_{min}} = 0.8$ are used to decide if R represents a stem. If R falls outside these thresholds then it is removed from $\{R\}$. Figure 5.20 presents $\{R\}$, after the non-stem point clouds have been pruned, with each constituent point cloud uniquely coloured.

Finally, a closing check is then performed to confirm that the point clouds in the set $\{R\}$ solely comprise of stems. This confirmation is undertaken through PCA, as described in section 5.2.6. An architectural assumption is made that stems exit the ground within 35° normal to the forest floor. Using equation 5.10, the angle between λ_1 derived from R and the vector normal to the ground derived from the DEM, can be calculated. If this angle exceeds the threshold of 35°, then R is removed from $\{R\}$. Figure 5.21 presents $\{R\}$ after this check has been performed and the non-stem point clouds pruned, with each constituent point cloud uniquely coloured.

Figure 5.20: The point clouds, each uniquely coloured, that comprise the set $\{R\}$, after pruning via RANSAC cylinder model fitting to represent the stems identified across P_s.

Figure 5.21: The point clouds, each uniquely coloured, that comprise the set $\{R\}$, after pruning via PCA to confirm they represent the stems identified across P_s.

5.3.5 Stem extraction

The point clouds $R \in \{R\}$ provide the location of each stem inside P_d. The cylinder model that has been fitted to each cloud also provides the radius and lean of each identified stem. The fifth step of *treeseg* is to use this information to extract a set of point clouds, $\{S\}$, from P_d, $\{S\} \subset P_d$, that represent each of the stems in their entirety.

This is demonstrated here on one arbitrarily selected stem, R.

The first step in this stem extraction is to extract a point cloud, F, that encompasses all the points from both the stem in question and the surrounding area. This is undertaken by filtering P_d via an infinitely long cylinder model, as described in section 5.2.7. This infinitely long cylinder model has the same cylinder centreline vector and centroid as R, but with a radius ten times larger than R. Figure 5.22a presents point cloud F.

The second step in the stem extraction is to remove those points from F that represent the forest floor. This is undertaken by extracting a small section of F between the first 0–1 m in the z-axis. A plane model is then fitted to this section via the RANSAC methods described in section 5.2.7, where $N = 1000$. The inliers of this plane represent the forest floor and are removed from F, as presented in figure 5.22b.

Following the removal of the forest floor, F is partitioned into a set of point clouds, $\{A\}$, $\{A\} \subset F$, each representing topologically-connected surfaces. For each point cloud in the set, $A \in \{A\}$, the topologically connected surfaces may be a single surface such as a stem, or a mix of inter-connected surfaces such as clumps of vegetation and branch. This partitioning is undertaken through Euclidean Clustering, as described in section 5.2.2. Here, $d_{cmax} = 9\overline{d_{NN}}(z)$ and $N_{min} = 10$. Figure 5.23a presents the set $\{A\}$ with each constituent point cloud uniquely coloured.

Next, each point cloud comprising $\{A\}$ is further partitioned into the set of point clouds, $\{B\}$, $\{B \subset \{A\}$, that solely represent topologically-connected single surfaces. This further partitioning is undertaken through region growing segmentation, as described in section 5.2.6. Here, $r_N = 9\overline{d_{NN}}(z)$, $c_{max} = 1$ and $\theta_{max} = 25°$. Figure 5.23b presents the set $\{B\}$ with each constituent point cloud uniquely coloured.

Following the partitioning of F, the next step is to determine which of point clouds in the set $\{B\}$ belong to the stem in question. This is achieved by creating a set of point clouds, $\{O\}$, that initially contain a single seed point cloud, O_{seed}. O_{seed} is one point cloud from $\{B\}$ that is known to belong to the stem. O_{seed} is readily identified in $\{B\}$ as the cloud with the lowest mean distance to R. O_{seed} is then removed from $\{B\}$ and inserted into $\{O\}$.

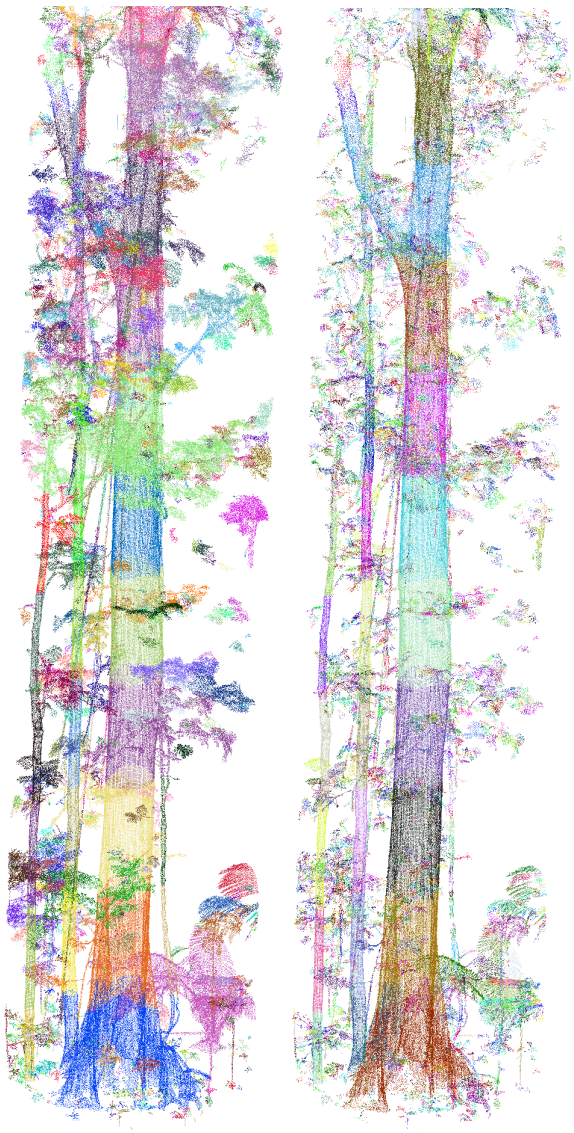
An iterative growing approach is then adopted to determine which point clouds in $\{B\}$ represent the part of the stem directly above and below O_{seed}. This determination is undertaken through PCA, as described in section 5.2.6.

(a) **(b)**

Figure 5.22: (a) For one arbitrarily selected stem that was identified in $\{R\}$, P_d is filtered via an infinitely long cylinder model driven by R to create the point cloud F that encompasses all the points from both the stem and surrounding area; and (b) points from the forest floor are removed from F by extracting the inlier points belonging to the RANSAC fitted plane model.

For the first iteration, three metrics are used decided if B, $B \in \{B\}$, belongs to the stem above and below O_{seed}. First, the smallest distance between O_{seed} and B, $d(O_{seed}, B)$. Second, the angle between the principal Eigenvectors, $\theta(\lambda_{1,O_{seed}}, \lambda_{1,B})$. And third, the ratio between the principal Eigenvalues, $v_{1,O_{seed}}$ and $v_{1,B}$.

For $B \in \{B\}$ these three metrics are computed. Each metric can then be normalised between 0 and 1. For example in the instance of the angle between the principal Eigen-



(a) **(b)**

Figure 5.23: (a) The point clouds, each uniquely coloured, that comprise the set $\{A\}$, partitioning F via Euclidean Clustering to represent topologically-connected surfaces; and (b) the point clouds, each uniquely coloured, that comprise the set $\{B\}$, further partitioning $\{A\}$ via region growing segmentation to represent topologically-connected single surfaces.

vectors, this normalisation would be performed as:

$$\theta_N = \frac{\theta(\lambda_{1,O_{seed}}, \lambda_{1,B}) - \theta_{min}}{\theta_{max} - \theta min} \tag{5.17}$$

Where θ_{min} and θ_{max} are the smallest and largest value of θ encountered across $B \in \{B\}$.

The average of these three normalised metrics for, μ, is defined as:

$$\mu = \frac{d_N + \theta_N + v_N}{3} \tag{5.18}$$

The two point clouds with the minimum values of μ are determined here as representing the portion of the stem above and below O_{seed}. As such, they are removed from $\{B\}$ and inserted into $\{O\}$. The two new constituents in $\{O\}$ act as new seed points for the second iteration. The process is repeated until either exhaustion of $B \in \{B\}$, or until none of B have a value of μ less than 0.5. Once complete, the set $\{O\}$ can be amalgamated into the single stem point cloud S. For the example stem demonstrated here, figure 5.24a presents S in black and R, from which O_{seed} was derived, overlaid in red.

It can be seen in figure 5.23a that S also contains a significant section of unwanted branching that belongs to the stem in question. This is because the point clouds in $\{B\}$ comprising this branching conformed to the above connectivity testing. This section of branching requires removal prior to the crown extraction in the next stage of *treeseg*. This removal is undertaken through RANSAC cylinder fitting, as described in section 5.2.7. From a starting point of 3 m up the stem in the z-axis, iteratively, 1.5 m slices of the stem are taken. Each slice is segmented into five equal sections by again filtering in the z-axis. For each of these sections and for the slice itself, a cylinder model is fitted using the RANSAC approach with $N = 1000$.

Two metrics are then used to determine if the point cloud from the slice solely belongs to the stem in question, or includes other elements such as branching. First, the coefficient of variation, CV, across the radii derived from the five sections. The second metric, the ratio between the radius of the cylinder model fitted to the slice and the mean radius from the five cylinder models fitted to the subsection, r_r. These two metrics encapsulate the stability of the fitted cylinder model, and as such can be interpreted as to how cylindrical the underlying surface is.

Here, the thresholds $CV_{max} = 0.05$ and $r_{r_{min}} = 0.9$ are used to make this decision. If the point cloud from this slice falls outside these thresholds then the iterations cease. When this occurs, the slice and all points laying above the slice are removed from S. Figure 5.24b presents S, after this removal process has been performed.

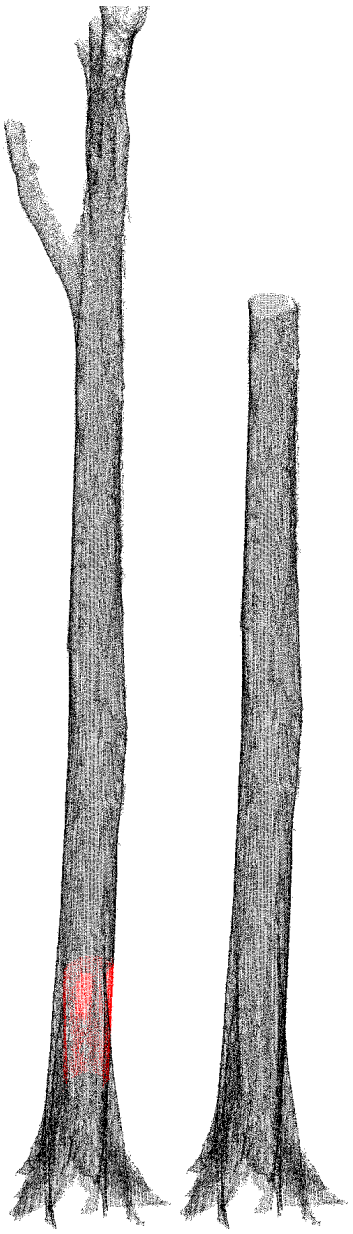


Figure 5.24: (a) The stem point cloud, S, coloured black, that has been iteratively grown from the seed point derived from R, coloured red; and (b) S after branching has been removed via RANSAC cylinder fitting.

5.3.6 Crown extraction

The point clouds in $\{S\}$ provide the extracted point clouds of each stem inside P_d up to the location of first branching. The sixth step of *treeseg* is to extract the crown belonging to each of the stems. This process is demonstrated here on the same example tree used in section 5.3.5.

The crown extraction begins with a cube point cloud, H, of edge length thirty times larger than the radius of the stem (derived from the cylinder model fitted to R) extracted

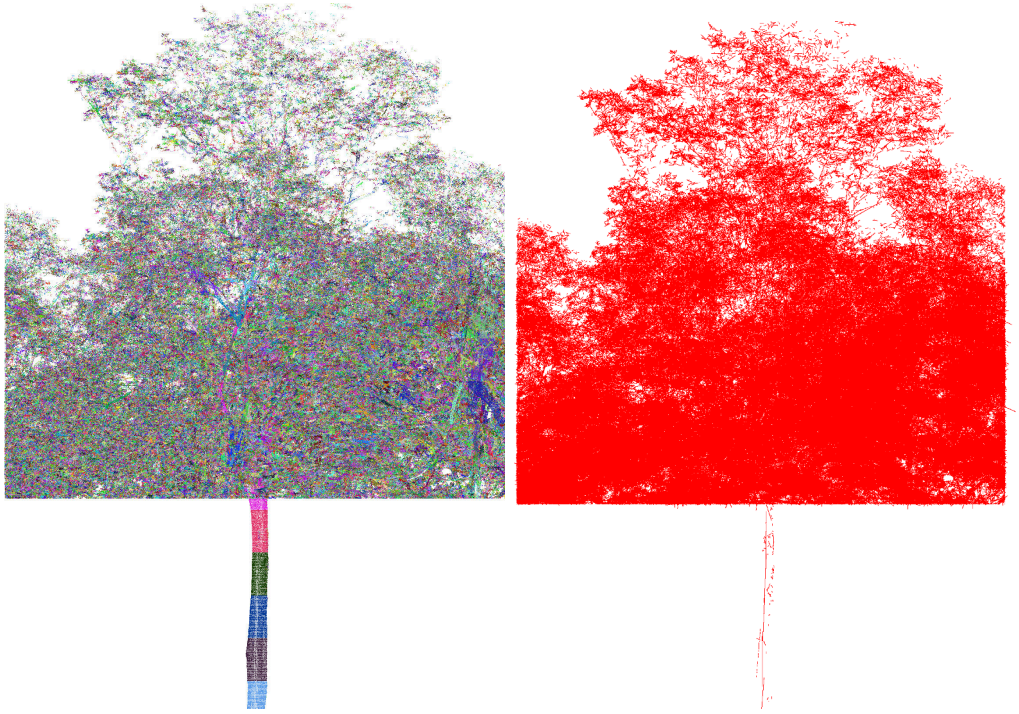
from P_d. The position of this cube in P_d is such that the centre of the bottom face is aligned with the top of S. H then, will contain the crown of S and (potentially) neighbouring canopy vegetation. H is then partitioned into a set of point clouds $\{G\}$, $\{G\} \subset H$, each representing topologically-connected single surfaces. This partitioning is undertaken through Euclidean clustering and region growing segmentation, as described in sections 5.2.2 and 5.2.6. Here, $d_{cmax} = 9\overline{d_{NN}}(z)$, $N_{min} = 1$, $r_N = 9\overline{d_{NN}}(z)$, $c_{max} = 1$ and $\theta_{max} = 25°$. Figure 5.25a presents the set $\{G\}$ with each constituent point cloud uniquely coloured.

The second step in the crown extraction is to decide which of the point clouds in the set $\{G\}$ belong to the tree in question. This is undertaken through PCA, as described in section 5.2.4. For $G \in \{G\}$, figure 5.25b presents the principal Eigenvector across the extent of the point cloud. A cross-section through the canopy has also been provided in figure 5.26.

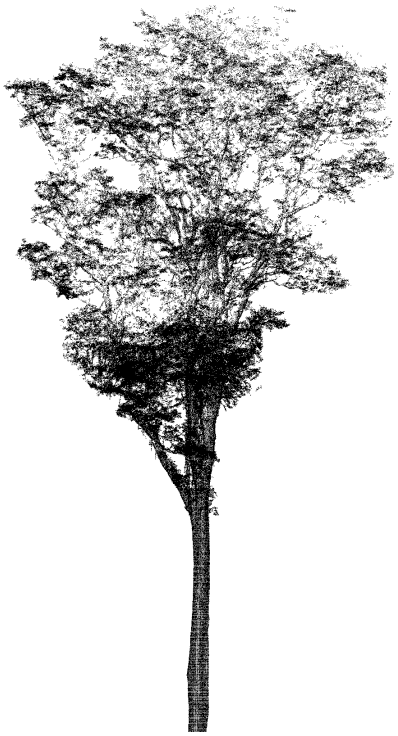
In a similar way to the stem extraction methods presented in section 5.3.5, an iterative growth approach is used to determine whether the point clouds in the set $\{G\}$, belong to the tree point cloud, T. T initially consists of just S. The seed point T_{seed} is obtained from the tip of S. The same three metrics are used to determine connectivity. They are, first, the smallest distance between T_{seed} and G, $d(T_{seed}, G)$. Second, the angle between the principal Eigenvectors, $\theta(\lambda_{1,T_{seed}}, \lambda_{1,G})$. And third, the ratio between the principal Eigenvalues, $v_{1,T_{seed}}$ and $v_{1,G}$.

As opposed to the stem extraction, where only a single point cloud above or below the seed point can belong to the stem, in the crown, a number of point clouds representing multiple branches must be permitted to be grown from a single seed. Instead of normalisation then, a number of hard thresholds are implemented. Here, $d_{max} = 2\overline{d_{NN}}(z)$, $\theta_{max} = 30°$ and $v_{max} = 50$. For $G \in \{G\}$, those point clouds conforming to these thresholds are considered part of the tree. As a result, they are appended to T and act as new seeds for the next iteration. The process is repeated until either exhaustion of $G \in \{G\}$, or until none of G in the current iteration conform to these thresholds.

A further constraint placed on the connectivity testing is that the principal Eigenvalue of the seed must be greater than or equal to half the size of the principal Eigenvalue of G. This constraint prevents the algorithm from growing into neighbouring trees. Figure 5.25c presents T.



(a) (b)



(c)

Figure 5.25: (a) The cube point cloud, H, is extracted from P_d such that it contains the crown of S and neighbouring canopy vegetation, here H has been partitioned into the uniquely coloured point clouds comprising the set $\{G\}$ through Euclidean clustering and region growing segmentation; (b) the principal Eigenvector derived via PCA for each point cloud in set $\{G\}$; and (c) those point clouds in $\{G\}$ that have been determined to belong to the tree in question via connectivity testing.

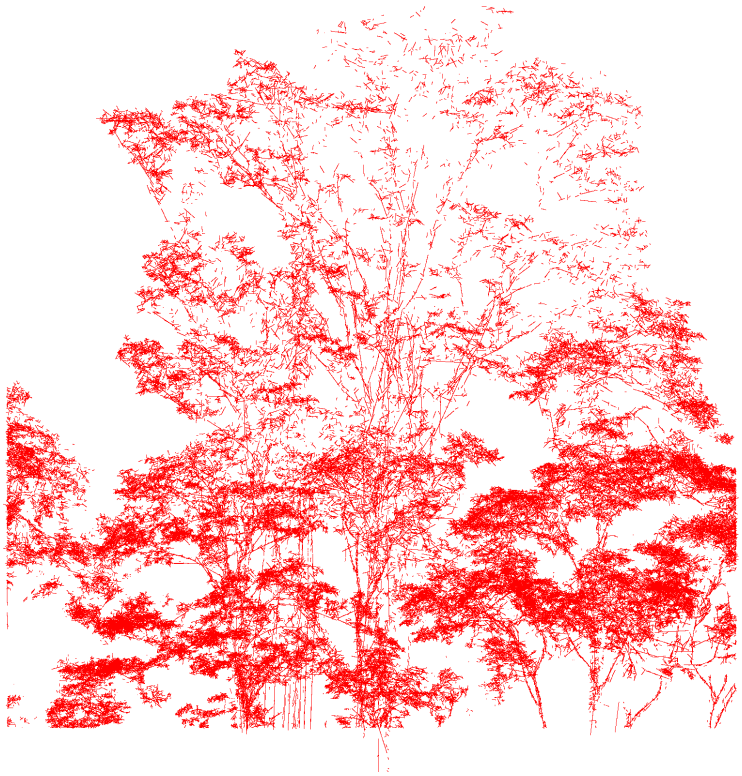


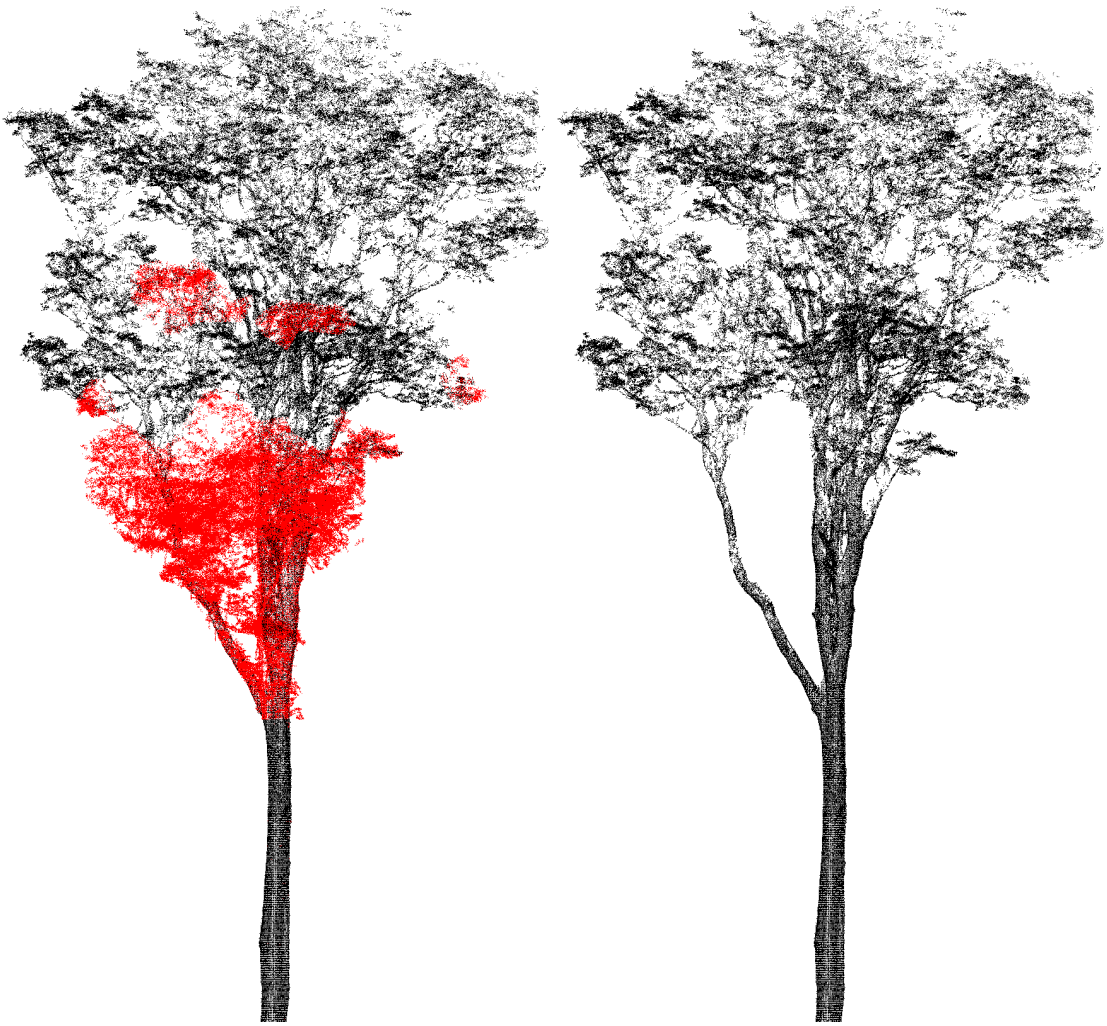
Figure 5.26: A slice through figure 5.25b to show a cross-section of the principal Eigenvectors in the canopy from the point cloud set $\{G\}$ derived via PCA.

5.3.7 Manual refinement

When the above steps have been completed, the point clouds in $\{T\}$ provide the extracted point clouds of each tree inside P_d in their entirety. The hard thresholds that have been employed in the connectivity testing during the crown extraction in section 5.3.6 are rather relaxed. That is, here, it was more desirable to segment additional elements of the canopy not belonging to the tree than to segment out elements of the crown belonging to the tree.

As a result of this conservative approach, this will occasionally require the seventh step of *treeseg*, the manual refinement of $\{T\}$. This is demonstrated here on the same example tree used in section 5.3.6. Figure 5.27a presents T that has been automatically extracted up to this point by steps 1-6 of *treeseg*. The points coloured black belong to the tree in question whilst the points coloured red belong to other vegetation in the canopy. The removal of these erroneous points, as shown in figure 5.27b, is undertaken through 'by hand extraction'.

In this particular instance, the cause of the inclusion of these erroneous points, and hence the requirement for this manual step, arose due to a smaller stem interacting with



(a) **(b)**

Figure 5.27: (a) The tree-level point cloud automatically extracted by steps 1-6 of *treeseg* with the points belonging to the tree coloured in black and the erroneously segmented points coloured in red; (b) the tree-level point cloud after manual refinement via 'by hand extraction'.

a branch of the tree in question. This interaction is illustrated in figure 5.28. During the connectivity testing, this smaller stem was determined to be part of the tree. This is because the small stem and branching are touching with the resultant angle between to the two being lower than the hard threshold.

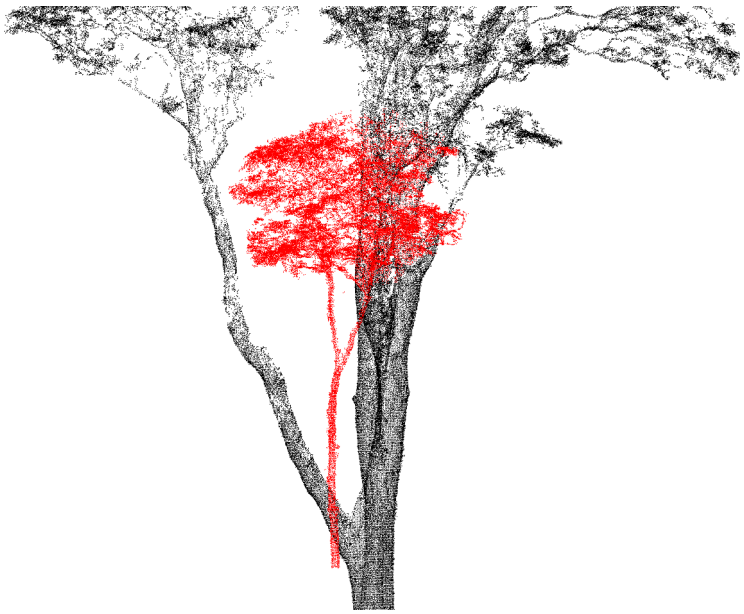


Figure 5.28: The small stem (red) interacting with the branching from the tree in question (black), that conformed to the hard thresholds during the connectivity testing that resulted in its inclusion during the crown extraction.

5.3.8 Retrieving tree-level parameters of structure

The tree-level point clouds in $\{T\}$ now explicitly represent the individuals extracted from P_d. The final eighth step of *treeseg* is to retrieve traditional tree-level parameters of structure. Considered here are the x and y coordinates, the bole diameter, D, tree height, H, and the crown extent, C_r. The retrieval of these parameters is demonstrated on the same example tree used in section 5.3.7.

The traditional measurement of D, obtained using a tape measure, is taken at a height of 1.3 m. However, if some buttress is present, then the measurement of D is taken at the lowest point at which the buttress will no longer impose on the measurement. To emulate this in the TLS data, an iterative approach is undertaken using RANSAC cylinder fitting, as described in section 5.2.7.

This iterative approach to estimate D begins with a slice, S of width 0.5 m taken through the z-axis of T such that the centre of the slice is at 1.3 m. Two additional slices of the same width are taken above and below, S_a and S_b such that their respective centres are at 0.8 m and 1.8 m. Across these three slices, a cylinder model is fitted using the RANSAC methods where $N = 1000$, providing the radii r_S, r_{S_a} and r_{S_b}. The change

between each radius, r_c is then calculated, for example as:

$$r_c(r_S, r_{S_a}) = \frac{r_S - r_{S_a}}{r_S} \tag{5.19}$$

The mean percentage difference, $\overline{r_c}$, is defined as:

$$\overline{r_c} = \frac{r_c(S, S_a) r_c(S, S_b)}{2} \tag{5.20}$$

$\overline{r_c}$ then, permits an assessment of how stable the cylinder model is across the local section of stem in question. If $\overline{r_c}$ is below 0.1 then the cylinder model is considered stable and D is estimated here as $D = 2r_s$. If not, then the slices are respectively raised by 0.1 m and the process is repeated until stability if found.

Figure 5.29 presents the first 5 m of the stem of T. In this figure, S, at which D was estimated is coloured red; the two slices, S_a and S_b are coloured green. It can be seen in this instance that the point of measurement occurs after the buttress that is present in T. Figure 5.30b presents a birds-eye view of S. Figure 5.30a then presents the slice at 1.3 m had this iterative approach not being implemented.

With the retrieval of D complete, the x and y coordinates can be trivially derived from the centroid of S through equation 5.5. Finally, H and C_r are trivially derived from the 3D bounding box of T. Figure 5.31a defines the lowest and highest value of $z_p \in T$, z_{min} and z_{max} respectively. H can then be estimated as:

$$H = z_{max} - z_{min} \tag{5.21}$$

Figure 5.31b defines the long and short lengths of the 2D (x-y) bounding box, C_A and C_B respectively. C_r can then be estimated as:

$$C_r = \sqrt{(C_A^2 + C_B^2)} \tag{5.22}$$

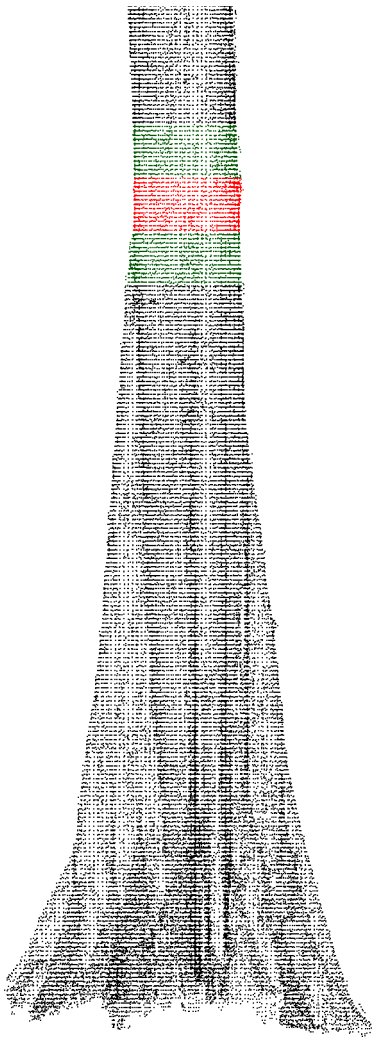


Figure 5.29: Estimation of D is made through an iterative RANSAC cylinder model fitting approach to avoid measurement across a buttress; the three slices S (red), S_a (green) and S_b (green) recursively climb the stem until stabilisation of the cylinder model occurs, as illustrated here.

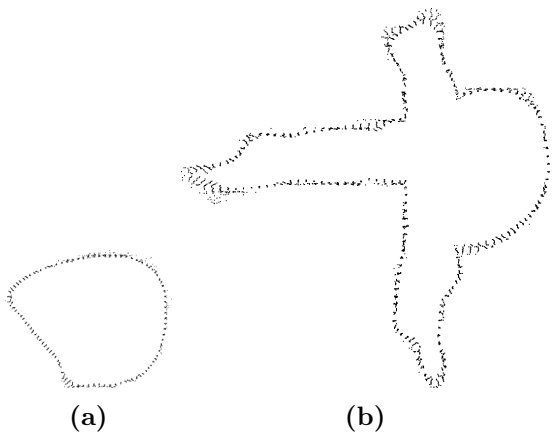


Figure 5.30: (a) Birds-eye view of S in figure 5.29; and (b) the birds-eye view of the stem at the traditional point of measurement at 1.3 m.

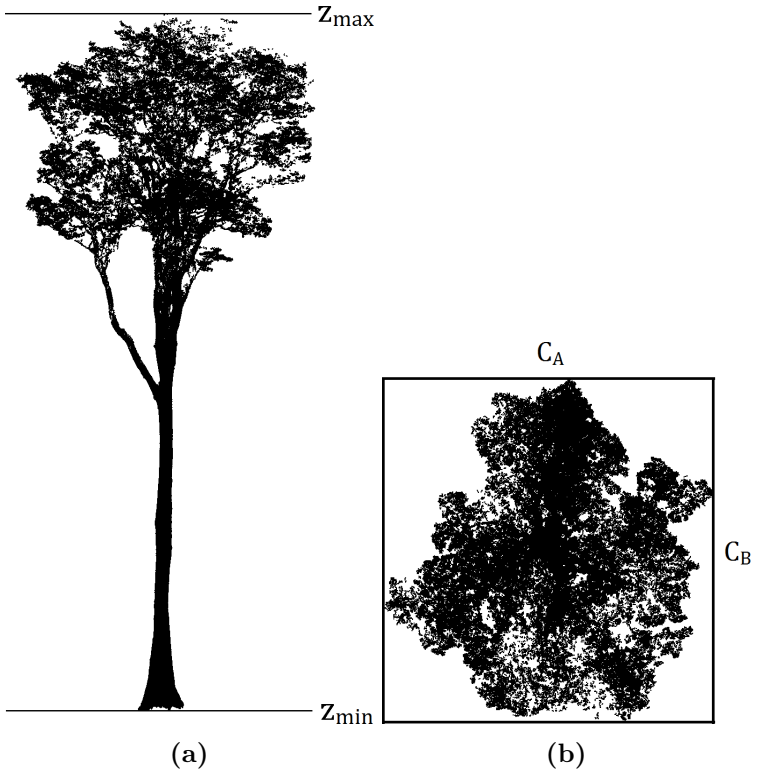


Figure 5.31: (a) The maximum and minimum point coordinates in the z-axis, z_{max} and z_{min}, from which H is estimated; and (b) the 2D (x-y) bounding box lengths, C_A and C_B from which C_r is estimated.

5.3.9 Pruning leaf returns

This section details a rudimentary method for pruning leaf returns from the set of tree-level point clouds $\{T\}$. This is not listed as a step of the *treeseg* algorithm, but is included here as it is of potential interest, as leaf material is central to tree ecophysiology applications. This method is demonstrated on a single crown point cloud, C, as seen in figure 5.33a.

The first step of this leaf pruning method is to partition C into the set of point clouds, $\{R\}$, $\{R\} \subset C$,that represent topologically-connected single surfaces. This partitioning is undertaken through region growing segmentation, as described in section 5.2.6. Here, $r_N = 9\overline{d_{NN}}(z)$, $c_{max} = 1$ and $\theta_{max} = 25°$. Figure 5.33a presents the set $\{R\}$ with each constituent point cloud uniquely coloured.

Following this PCA is performed on each point cloud in the set $\{R\}$, as described in section 5.2.6. As was also described in section 5.2.6, and can be seen in figure 5.33a, the point clouds of $\{R\}$ from the branching appear clearly structured. Towards the high order branching, where the point clouds of $\{R\}$ incorporate points from leafy material or

are entirely composed of points from leafy material, this apparent structure disappears. This is because the underlying surface, the leafy material, is far more randomly orientated and distributed than the woody components. As a result, they appear more 'fuzzy' than their woody material counterparts. This method attempts to exploit this 'fuzziness' through classification.

This classification is undertaken here using Sneed and Folk Form Indices (Sneed et al. 1958) to categorise how platy, bladed, elongated or compact the shape of an object is. As illustrated in figure 5.32, this broad categorisation is achieved using three metrics based on the dimensions of an object. Here, in lieu of knowing the underlying dimensions of the surface that R represents, the Eigenvalues v_1, v_2 and v_3 are used to represent the magnitude of the longest, intermediate and shortest axes of R. The first two metrics, A and B, are ratios of the Eigenvalues, defined as:

$$A = \frac{v_3}{v_1} \tag{5.23}$$

$$B = \frac{v_1 - v_2}{v_1 - v_3} \tag{5.24}$$

The third metric, Ψ, is the maximum projection sphericity, defined as:

$$\Psi = \sqrt[3]{\frac{v_3^2}{v_1 v_2}} \tag{5.25}$$

This classification has been undertaken here on $\{R\}$ using thresholds of $\Psi_{max} = 0.33$, $A_{max} = 0.2$, $B_{min} = 0.8$. For $R \in \{R\}$, those clouds falling inside these thresholds are placed into the set of wood point clouds, $\{W\}$, $\{W\} \subset \{R\}$. For $R \in \{R\}$, those clouds falling outside these thresholds are placed into the set of leaf point clouds, $\{L\}$, $\{L\} \subset \{R\}$. Figures 5.33b and 5.33c present $\{W\}$ and $\{L\}$ respectively. Figure 5.33d then presents $\{W\}$ and $\{L\}$ overlaid each other with the wood point clouds coloured brown and the leaf point clouds coloured green. It can be seen from these figures that a reasonable segmentation of the points has been performed. Although, as most prominently observed in figure 5.33b, there is some misclassification of points.

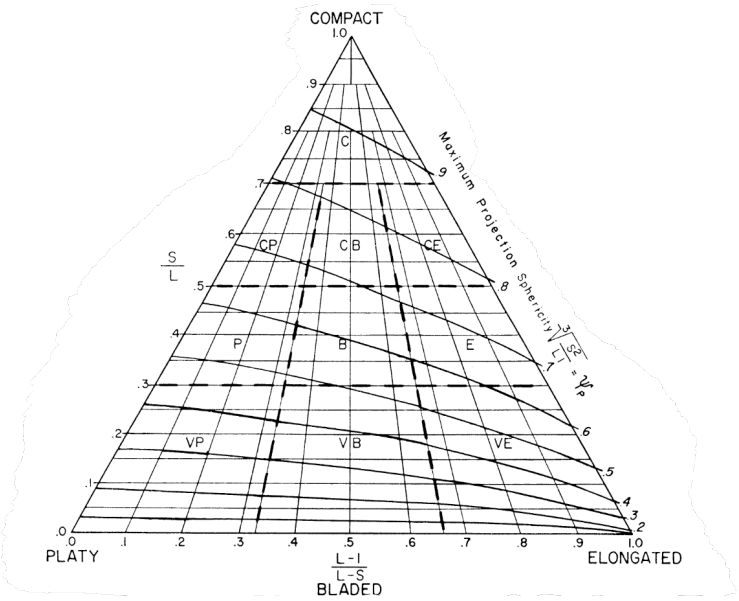


Figure 5.32: Sneed and Folk Form Indices used to classify objects by their shape into the categories platy, bladed, elongated or compact (Source: Sneed et al. 1958).

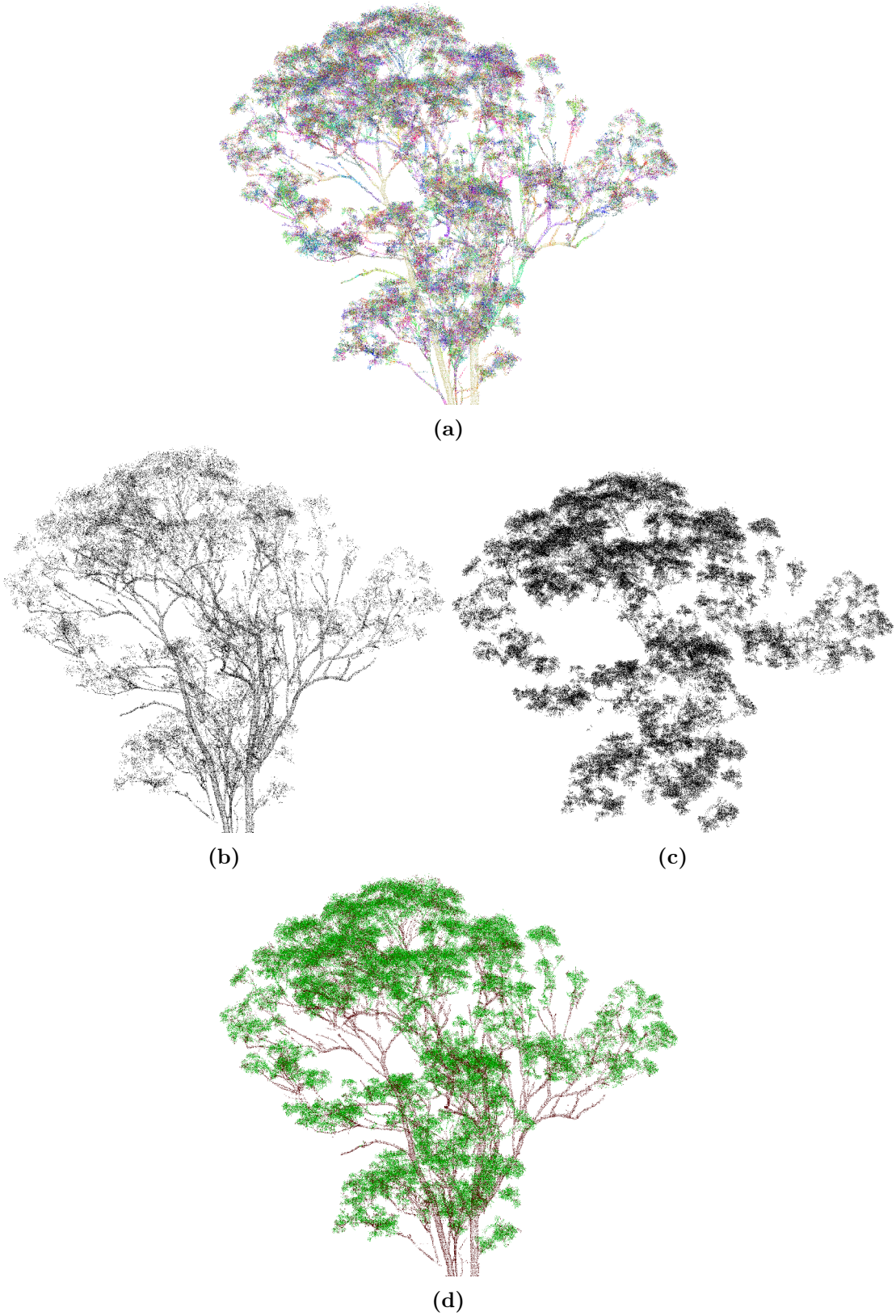


Figure 5.33: (a) Point cloud from a tree crown that has been partitioned into the set of point clouds $\{R\}$ via region growing segmentation; (b) and (c) the point clouds comprising the set of wood point clouds $\{W\}$ and leaf point clouds $\{L\}$ respectively, as classified via their Sneed and Folk Form Indices; and (d) the wood point clouds coloured brown and the leaf point clouds coloured green.

5.4 Results

In this section the results are presented from the application of *treeseg* to the plot-level point clouds obtained across the 3.25 ha of predominately tropical forest described in chapter 4 section 4.2. From these data, an attempt has been made to automatically extract all the stems inside the forest plots with $D \geq 0.2$ m.

The following sections present the results from the four individual forest plots. Each section begins with a table comparing the field data to the TLS data. Following this, a plot-scale illustration of each extracted tree-level point cloud is presented. Next, the stem map from the field data is overlaid on the stem map derived from the TLS data. Finally, a comparison of the distribution of stems by D and, where available H, between the field data and TLS data is provided.

5.4.1 Karawatha Forest Park, Australia (KARA-001)

As described in chapter 4 section 4.2.1, TLS data were acquired from the 0.25 ha censused *Eucalyptus* spp. open forest plot KARA-001 in Karawatha Forest Park, Australia. Applying the methods of *treeseg* to the plot-level point cloud of KARA-001, 28 stems were automatically extracted, of which 2 required manual refinement.

Table 5.1 compares the stem count and basal area between the field data and TLS data. It total, 5 additional individuals were extracted from the TLS data that were not present in the field data, representing a 34.2 % larger basal area over the plot.

	Field data	TLS data	% Difference
Stem Count	23	28	19.6
Basal Area ($m^2\,ha^{-1}$)	2.28	3.22	34.2

Table 5.1: Comparison of stem count and basal area between the field data and TLS data for plot KARA-001.

Figure 5.34 illustrates each of the 28 tree-level point clouds that have been extracted from the plot-level point cloud of KARA-001. Figure 5.35 then shows the stem map from the field data overlaid on the stem map derived from the TLS data. At the centre of the plot there appears to be some commonality between the two stem maps, although this commonality dissipates towards the plot boundaries.

Figure 5.36 presents the distribution of stems by D, using 0.1 m bins, across both the field and TLS data. In comparison to the field data, it can be seen that there are 2

fewer stems present in the TLS data across the 0.2–0.3 m range. Above this range, there are more stems present in the TLS data across all bins.

Figure 5.34: The 28 individual tree-level point clouds that have been extracted from the plot-level point cloud of KARA-001 via the *treeseg* algorithm.

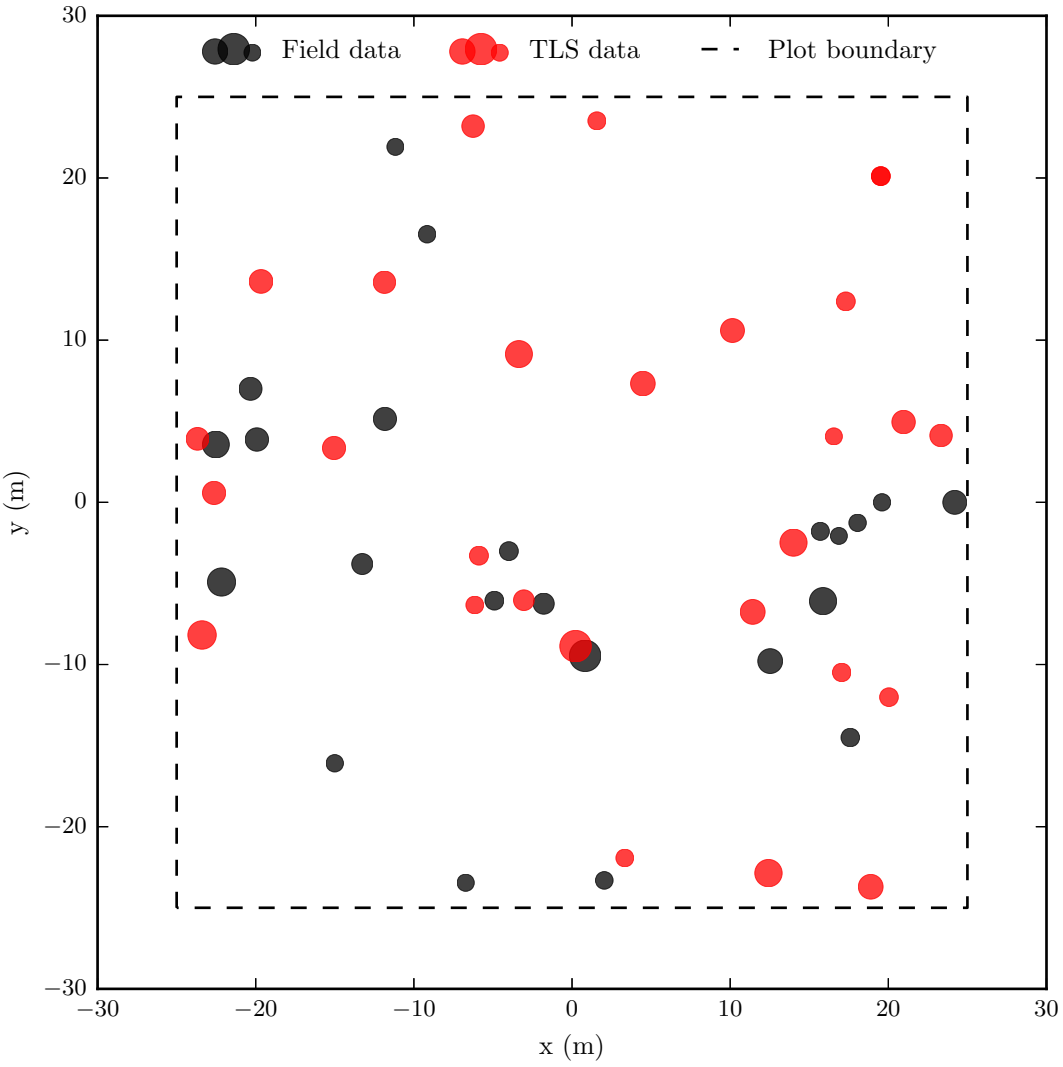


Figure 5.35: Stem map of KARA-001 overlaid on the stem map derived from the TLS data.

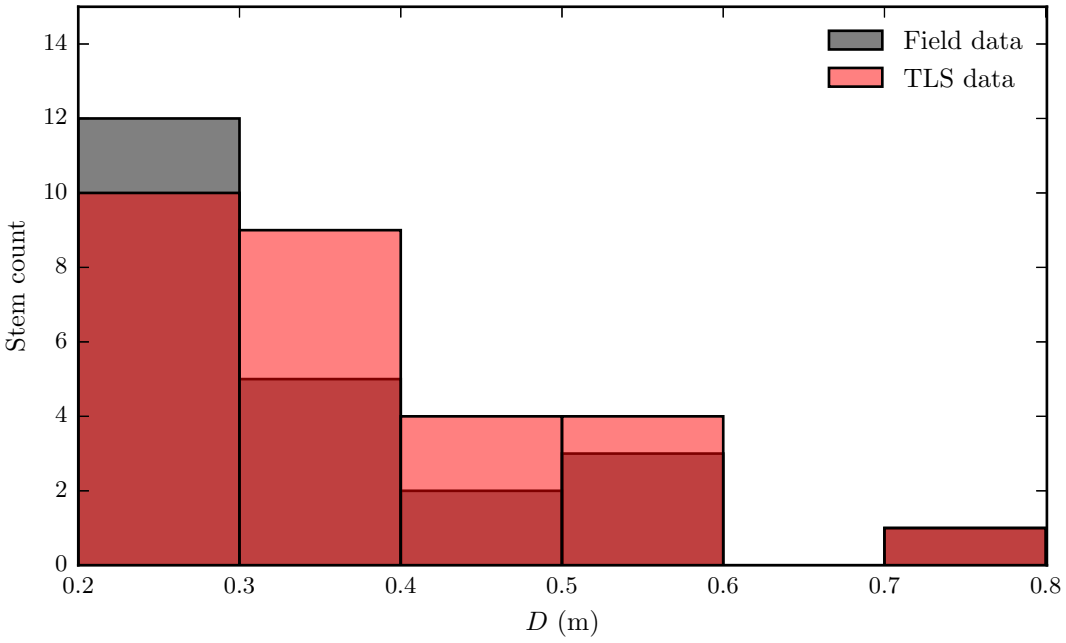


Figure 5.36: Distribution of stems by D across KARA-001 for both the field and TLS data.

5.4.2 Lopé National Park, Gabon (LPG-01)

As described in chapter 4 section 4.2.2, TLS data were acquired from the 1 ha censused moist, Terra Firme, lowland, mixed species, old-growth tropical forest plot LPG-01 in Lopé National Park, Gabon. Applying the methods of *treeseg* to the plot-level point cloud of LPG-01, 107 stems were automatically extracted, of which 19 required manual refinement.

Table 5.2 compares the stem count and basal area between the field data and TLS data. In total, 32 fewer individuals were extracted from the TLS data than are indicated to be present in the field data. Despite this, the basal area derived from these TLS data was 13.6 % larger than in the field data.

	Field data	TLS data	% Difference
Stem Count	139	107	26.0
Basal Area ($m^2\,ha^{-1}$)	28.46	32.62	13.6

Table 5.2: Comparison of stem count and basal area between the field data and TLS data for plot LPG-01.

Figure 5.37 illustrates each of the 107 tree-level point clouds that have been extracted from the plot-level point cloud of LPG-01. Figure 5.38 then shows the stem map from the field data overlaid on the stem map derived from the TLS data. It can be seen that across the full extent of the plot there is some commonality between the two stem maps, particularly in regard to the larger stems.

Figure 5.39 presents the distribution of stems by D, using 0.1 m bins, across both the field and TLS data. In comparison to the field data, across the 0.2–0.3 m range, there are approximately 60 % fewer stems present in the TLS data. Above $D = 0.4$ m there are generally more stems present in the TLS data than in the field data. Above $D = 1.0$ m in particular, there are several large individuals present in the TLS data that do not appear in the field data.

Figure 5.40 presents the distribution of stems by H, using 5 m bins, across both the field and TLS data. It can be seen across the 10–15 m range there are no individuals present in the TLS data. The distribution of stems across the remaining range varies significantly between the field and TLS data. Between the ranges of 30–40 m there are approximately 20 fewer individuals in the field data than in the TLS data. There is also significant contradiction between the field data and TLS data over the maximum tree

height, coming in at 45 m and 70 m respectively.

Figure 5.37: The 107 individual tree-level point clouds that have been extracted from the plot-level point cloud of LPG-01 via the *treeseg* algorithm.

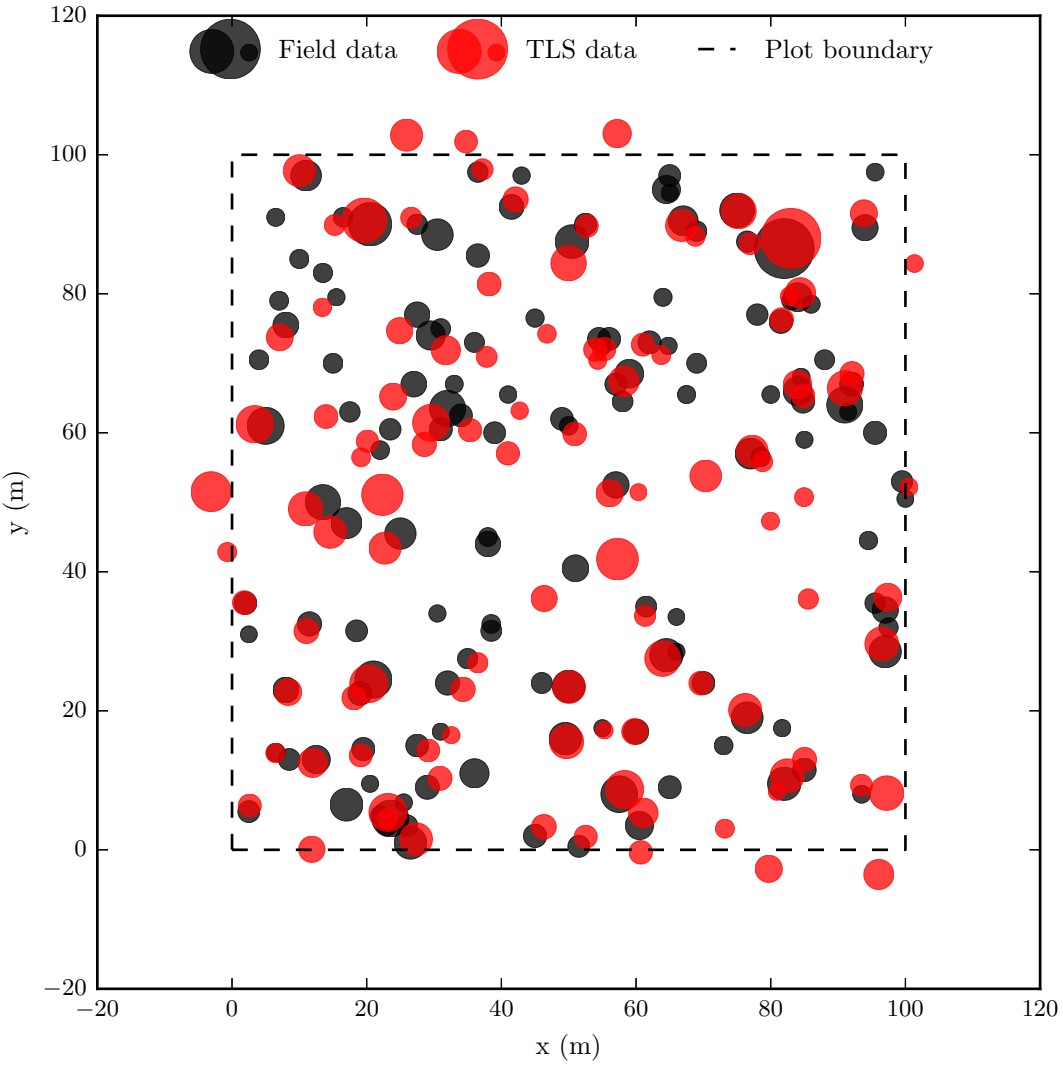


Figure 5.38: Stem map of LPG-01 overlaid on the stem map derived from the TLS data.

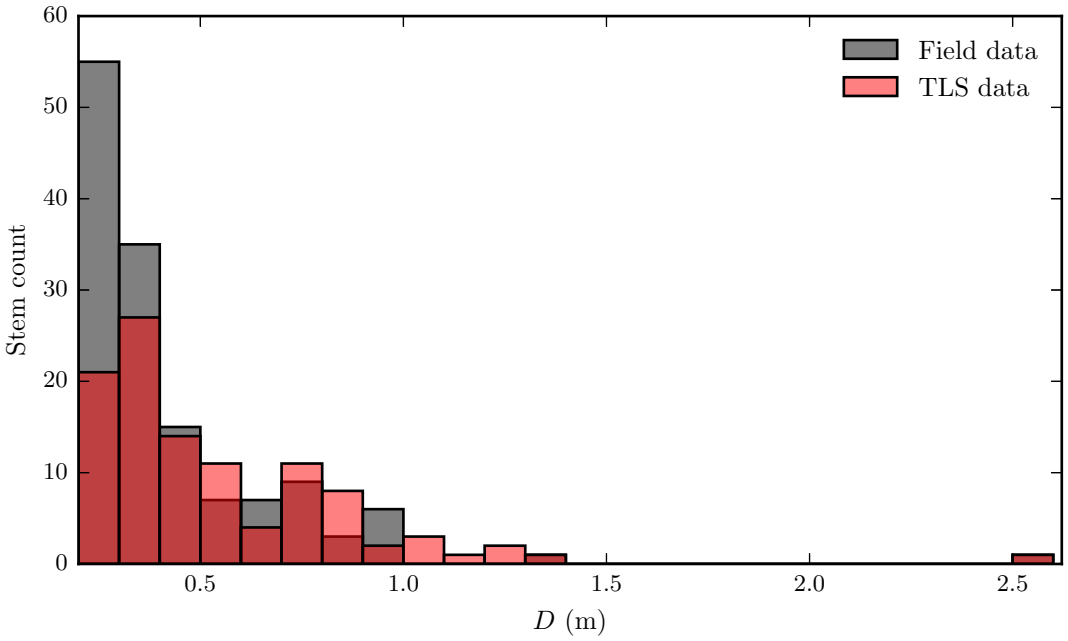


Figure 5.39: Distribution of stems by D across LPG-01 for both the field and TLS data.

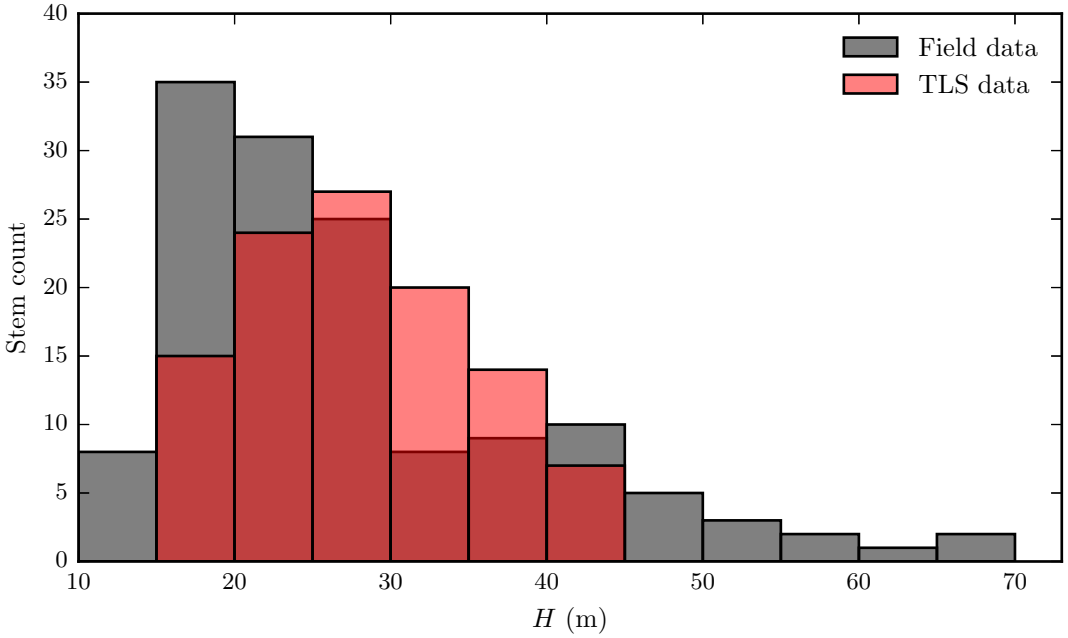


Figure 5.40: Distribution of stems by H across LPG-01 for both the field and TLS data.

5.4.3 Caixuanã National Forest, Brazil (CAX-A)

As described in chapter 4 section 4.2.4, TLS data were acquired from the 1 ha censused moist, Terra Firme, lowland, mixed species, old-growth tropical forest plot CAX-A in Caixuanã National Forest, Brazil. Applying the methods of *treeseg* to the plot-level point cloud of CAX-A, 151 stems were automatically extracted, of which 33 required manual refinement.

Table 5.3 compares the stem count and basal area between the field data and TLS data. It total, 63 fewer individuals were extracted from the TLS data than are indicated to be present in the field data. This resulted in a 12.1 % lower basal area across the plot.

	Field data	TLS data	% Difference
Stem Count	214	151	34.5
Basal Area ($m^2\,ha^{-1}$)	29.30	25.97	12.1

Table 5.3: Comparison of stem count and basal area between the field data and TLS data for plot CAX-A.

Figure 5.41 illustrates each of the 151 tree-level point clouds that have been extracted from the plot-level point cloud of CAX-A. Figure 5.42 then shows the stem map from the field data overlaid on the stem map derived from the TLS data. It can be seen across the plot that there appears to be almost no commonality between the two stem maps.

Figure 5.43 presents the distribution of stems by D, using 0.1 m bins, across both the field and TLS data. In comparison to the field data, across the 0.2–0.3 m range, there are approximately 50 % fewer stems present in the TLS data. Above $D = 0.4$ m the distribution is broadly similar although there is the noticeable absence of two stems at 1.2 m and 1.4 m in the TLS data.

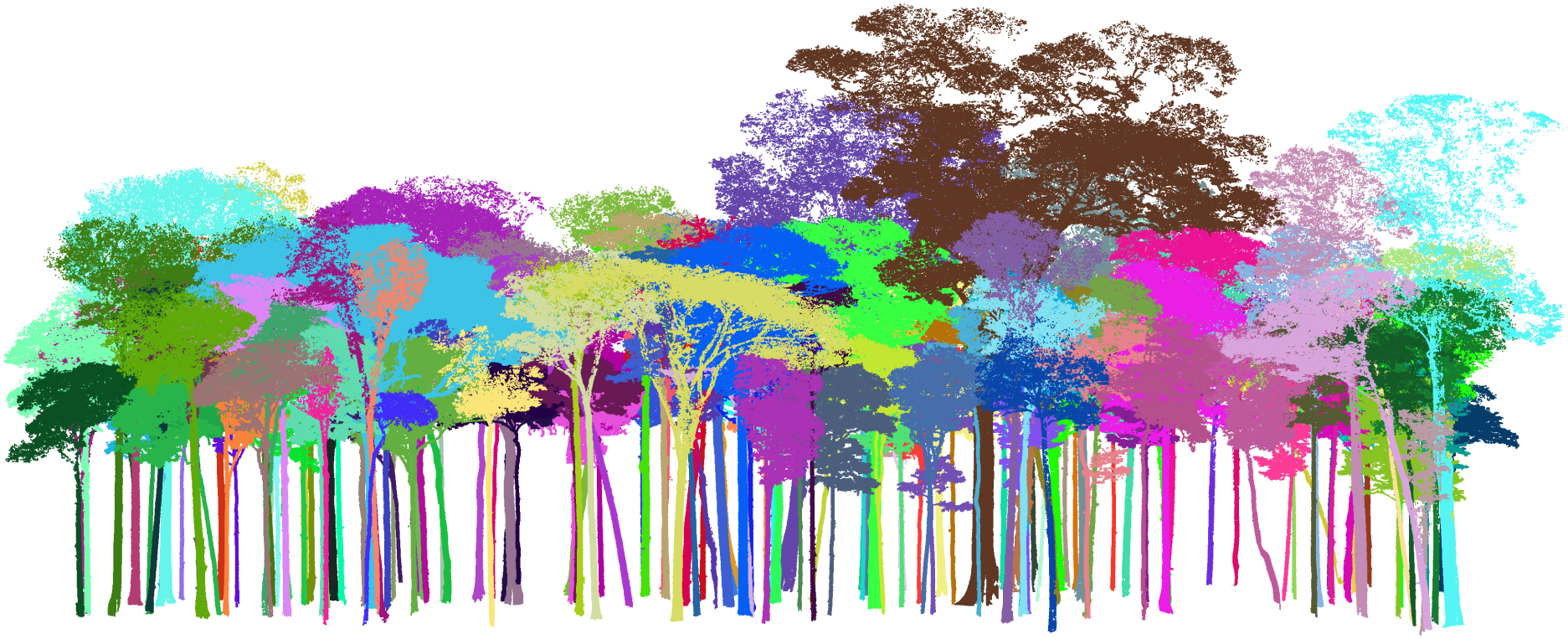


Figure 5.41: The 151 individual tree-level point clouds that have been extracted from the plot-level point cloud of CAX-A via the *treeseg* algorithm.

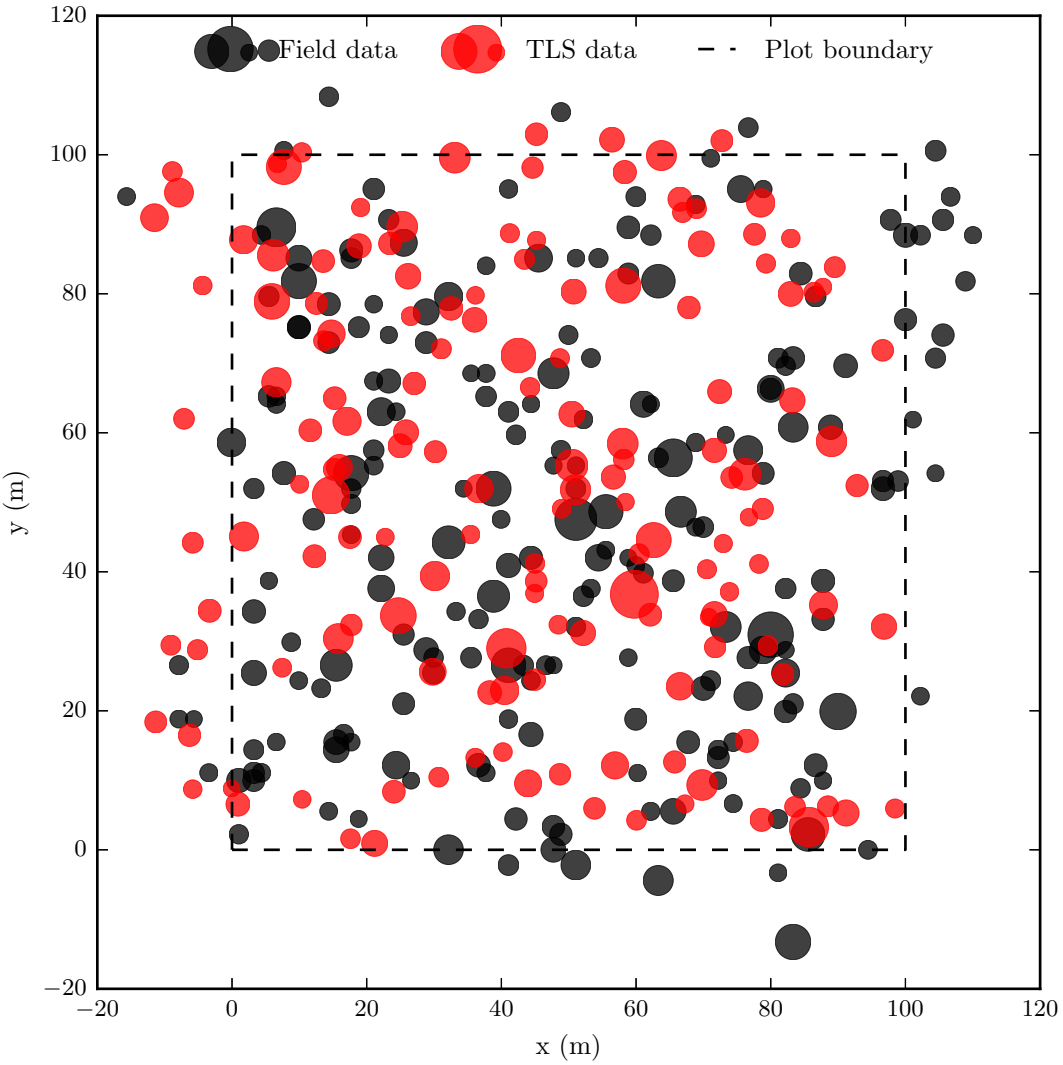


Figure 5.42: Stem map of CAX-A overlaid on the stem map derived from the TLS data.

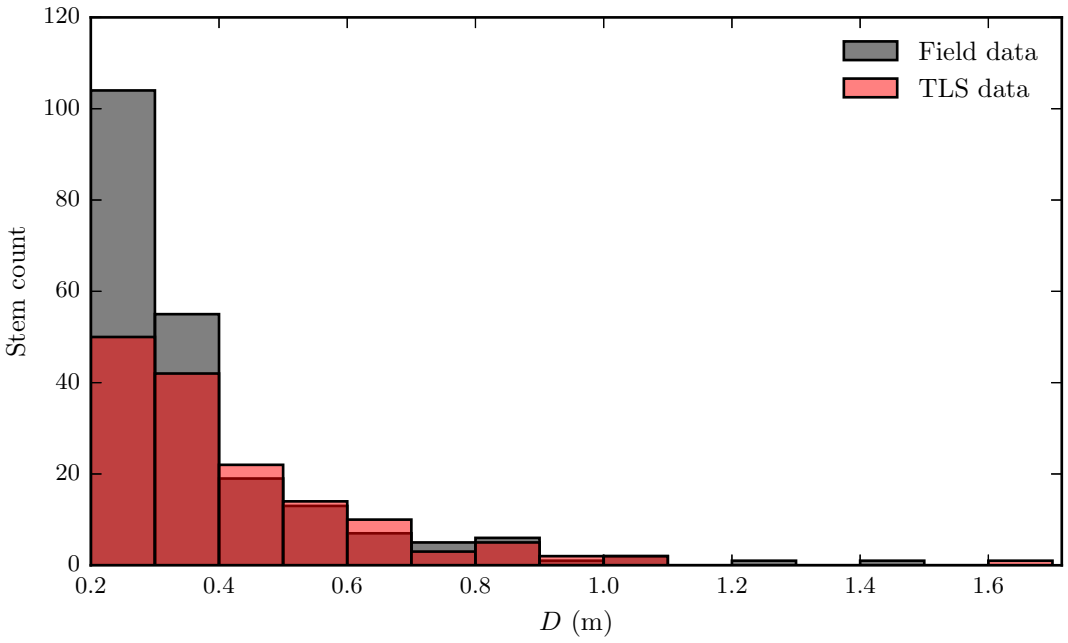


Figure 5.43: Distribution of stems by D across CAX-A for both the field data and TLS data.

5.4.4 Nouragues Nature Reserve, French Guiana (NOU-11)

As described in chapter 4 section 4.2.5, TLS data were acquired from the 1 ha censused moist, Terra Firme, lowland, mixed species, old-growth tropical forest plot NOU-11 in Nouragues Nature Reserve, French Guiana. Applying the methods of *treeseg* to the plot-level point cloud of NOU-11, 155 stems were automatically extracted, of which 45 required manual refinement.

Table 5.4 compares the stem count and basal area between the field data and TLS data. In total, 55 fewer individuals were extracted from the TLS data than are indicated to be present in the field data. This resulted in a 6.1 % lower basal area across the plot.

	Field data	TLS data	% Difference
Stem Count	210	155	30.1
Basal Area ($m^2\,ha^{-1}$)	29.59	27.83	6.1

Table 5.4: Comparison of stem count and basal area between the field data and TLS data for plot NOU-11.

Figure 5.44 illustrates each of the 151 tree-level point clouds that have been extracted from the plot-level point cloud of NOU-11. Figure 5.45 then shows the stem map from the field data overlaid on the stem map derived from the TLS data. It can be seen that across the extent of the plot there appears to be commonality between the stem maps. It can be seen that this commonality is stronger across the first half of the plot (in the x-axis), than in the second half of the plot.

Figure 5.46 presents the distribution of stems by D, using 0.1 m bins, across both the field and TLS data. In comparison to the field data, across the 0.2–0.1 m range, there are again approximately 50 % fewer stems present in the TLS data. Above 0.4 m the absolute stem count and distribution are broadly consistent.

Figure 5.47 presents the distribution of stems by H, using 5 m bins, across both the field and TLS data. It can be seen across the 10–25 m range there are significantly fewer individuals present in the TLS data than in the field data. Above 30 m, more individuals are present in the TLS data than in the field data. Both distributions appear Gaussian but with quite divergent means.

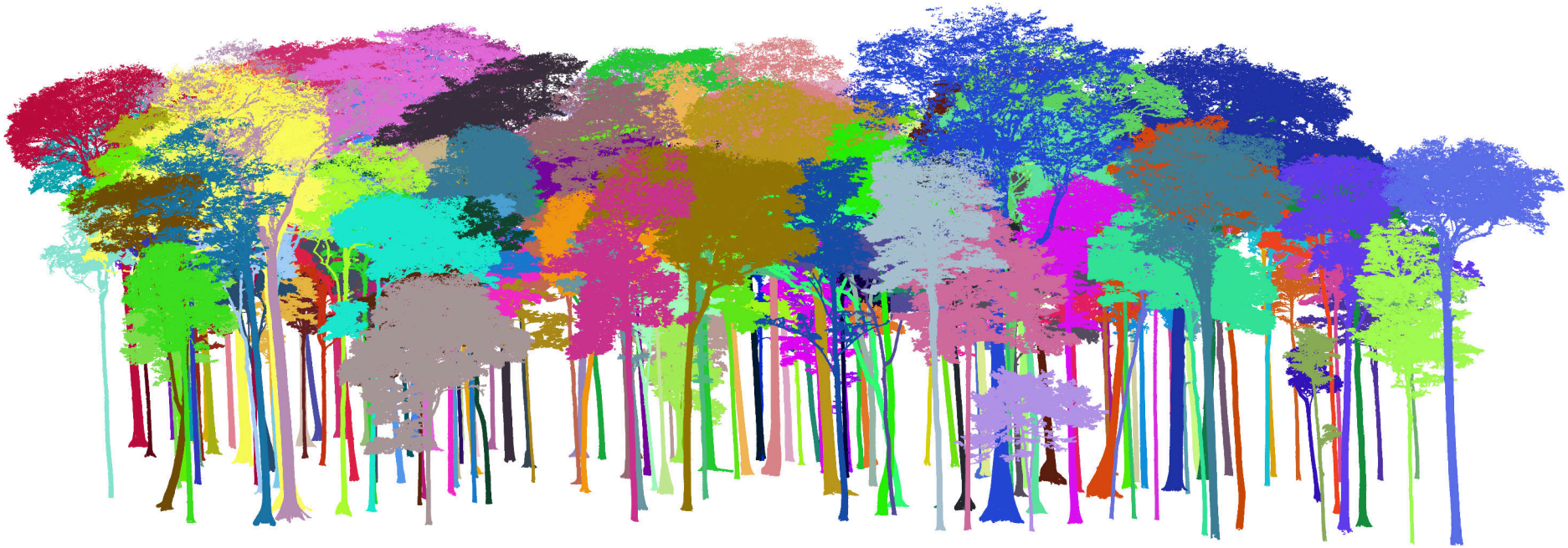


Figure 5.44: The 155 individual tree-level point clouds that have been extracted from the plot-level point cloud of NOU-11 via the *treeseg* algorithm.

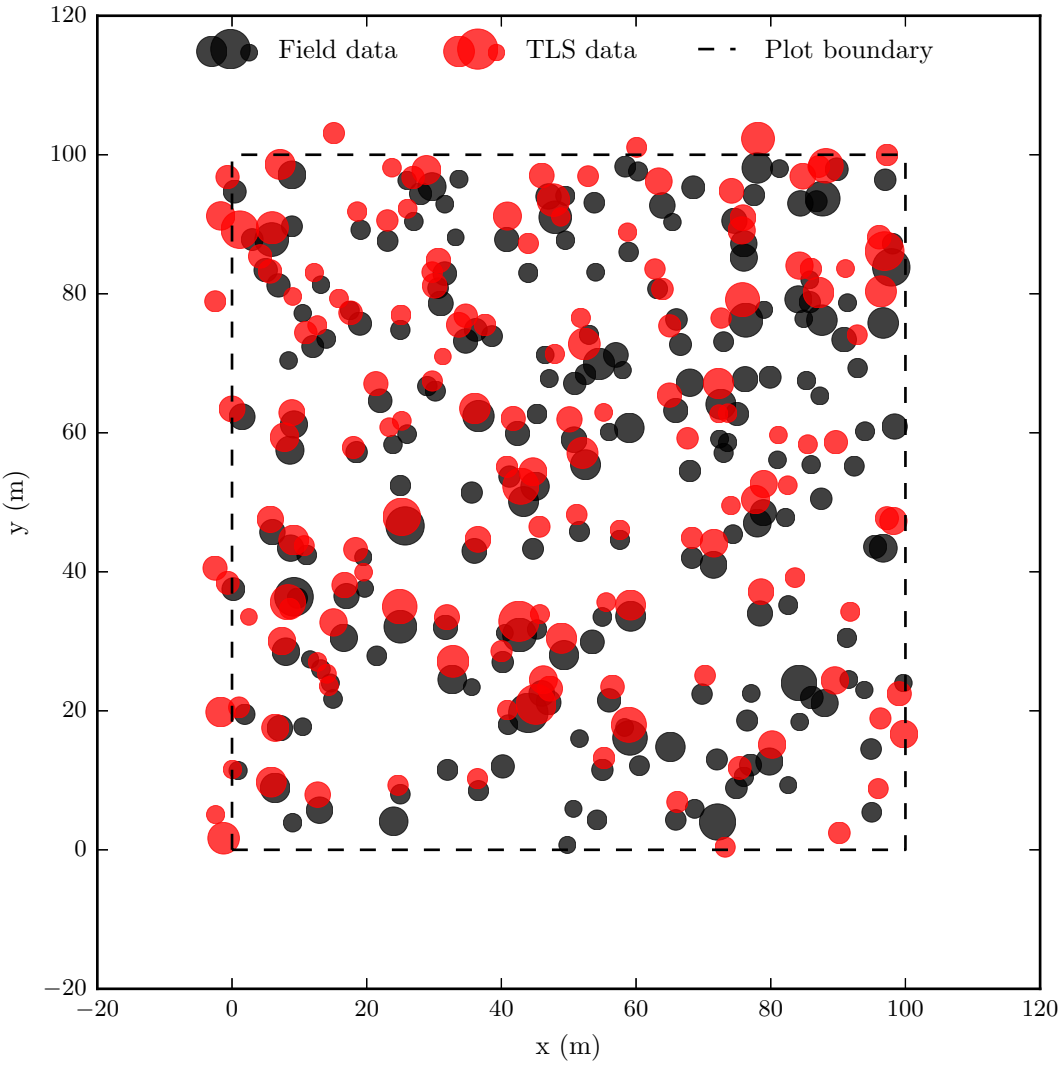


Figure 5.45: Stem map of NOU-11 overlaid on the stem map derived from the TLS data.

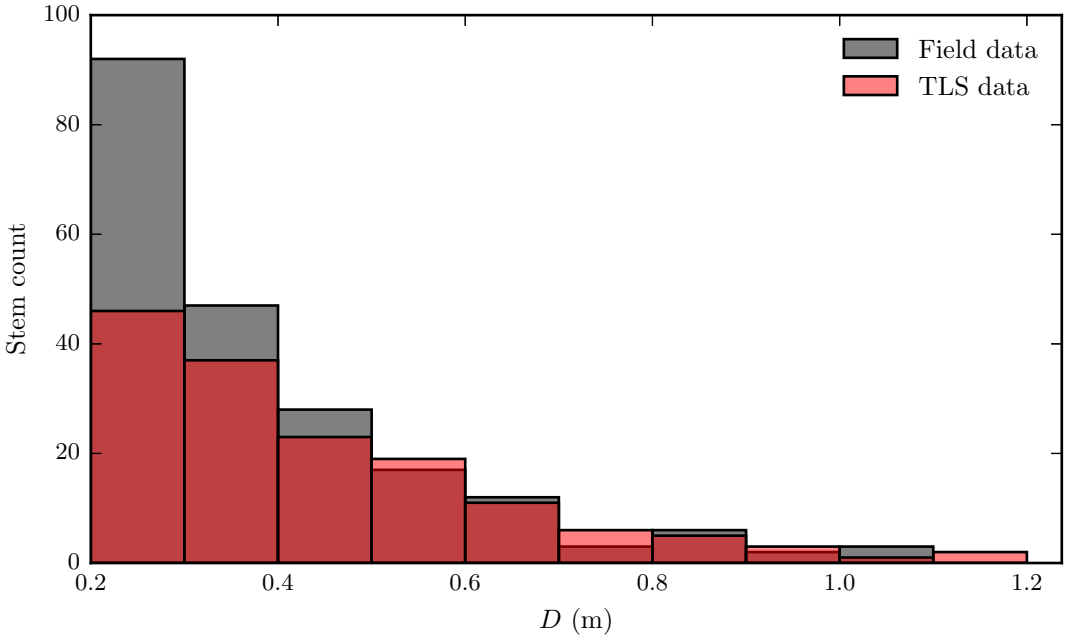


Figure 5.46: Distribution of stems by D across NOU-11 for both the field data and TLS data.

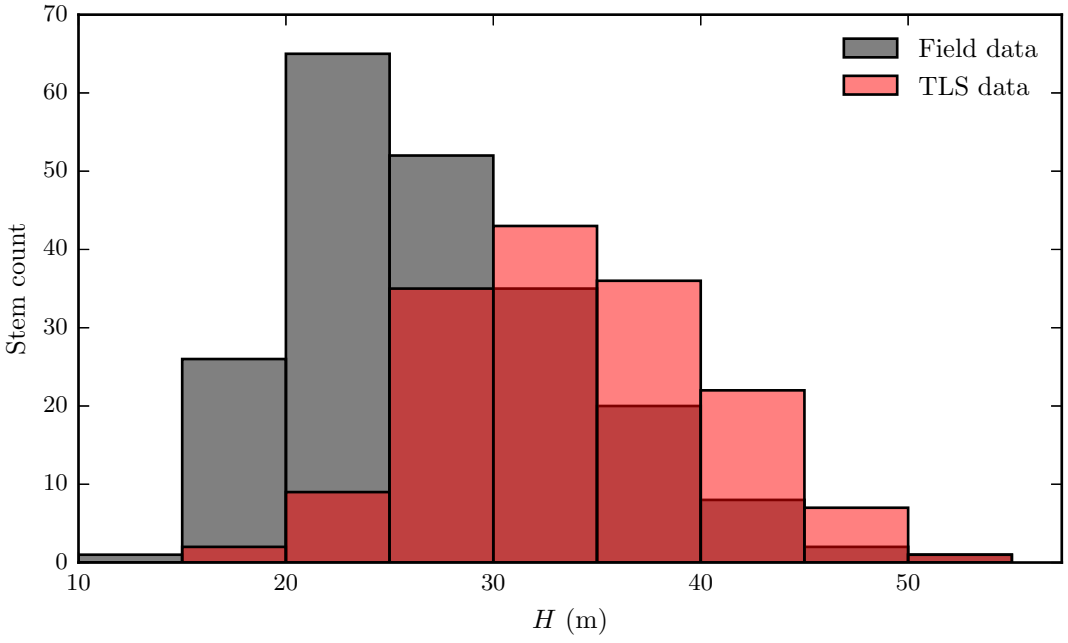


Figure 5.47: Distribution of stems by H across NOU-11 for both the field data and TLS data.

5.4.5 Tree-level parameters of structure

The previous four sections have each provided a comparison of the stem distribution by D and H across both the field data and TLS data for each individual forest plot. To demonstrate these data are indeed comparable, this section briefly presents results from an assessment of the goodness of the estimates of D and H that have been retrieved from the tree-level point clouds.

As described in section 5.3.8, D has been estimated here for each of the tree-level point clouds via an iterative RANSAC cylinder fitting routine. This method was proposed to avoid measurement of D across the buttresses of stems. To assess the accuracy of this method, 75 of the extracted tree-level point clouds were selected at random. For these 75 individuals, D was estimated via the methods of section 5.3.8 to provide D_A. D was also manually estimated to provide D_M. This manual estimation was achieved by extracting a small slice of the stem from each tree-level point cloud, either at the traditional point of measurement at 1.3 m, or, at the first position along the stem at which the buttress would not impose on the measurement of D. An example of such a slice can be seen in figure 5.30b. D_M then, was estimated by averaging the diameter of the major and minor axes of this slice via the Euclidean distance.

Figure 5.48 compares the estimates of D_A against D_M for the 75 randomly selected individuals. It can be seen that the estimates of D from the two independent methods are largely consistent with each other. The mean percentage difference between D_M and D_A is $-3.6\,\%$ with a standard deviation of $5.7\,\%$. The percentage difference across the two measurements ranges from -14.7–$11.2\,\%$. This indicates that there is a systematic tendency for the automatic method to come in slightly below the manual method. It is worth noting however that D_M is not equivalent to the measurement of circumference as would be made in the field. Therefore, D_M cannot be interpreted as validation data, but does provide some insight as to the goodness of D retrieved from the automatic methods.

With respect to H, there is little concern with the accuracy in the retrieval of this parameter from the tree-level point clouds. This is because each extracted tree-level point cloud was assessed to identify whether manual refinement was required. This assessment involved overlaying the tree-level point cloud with the surrounding canopy point cloud;

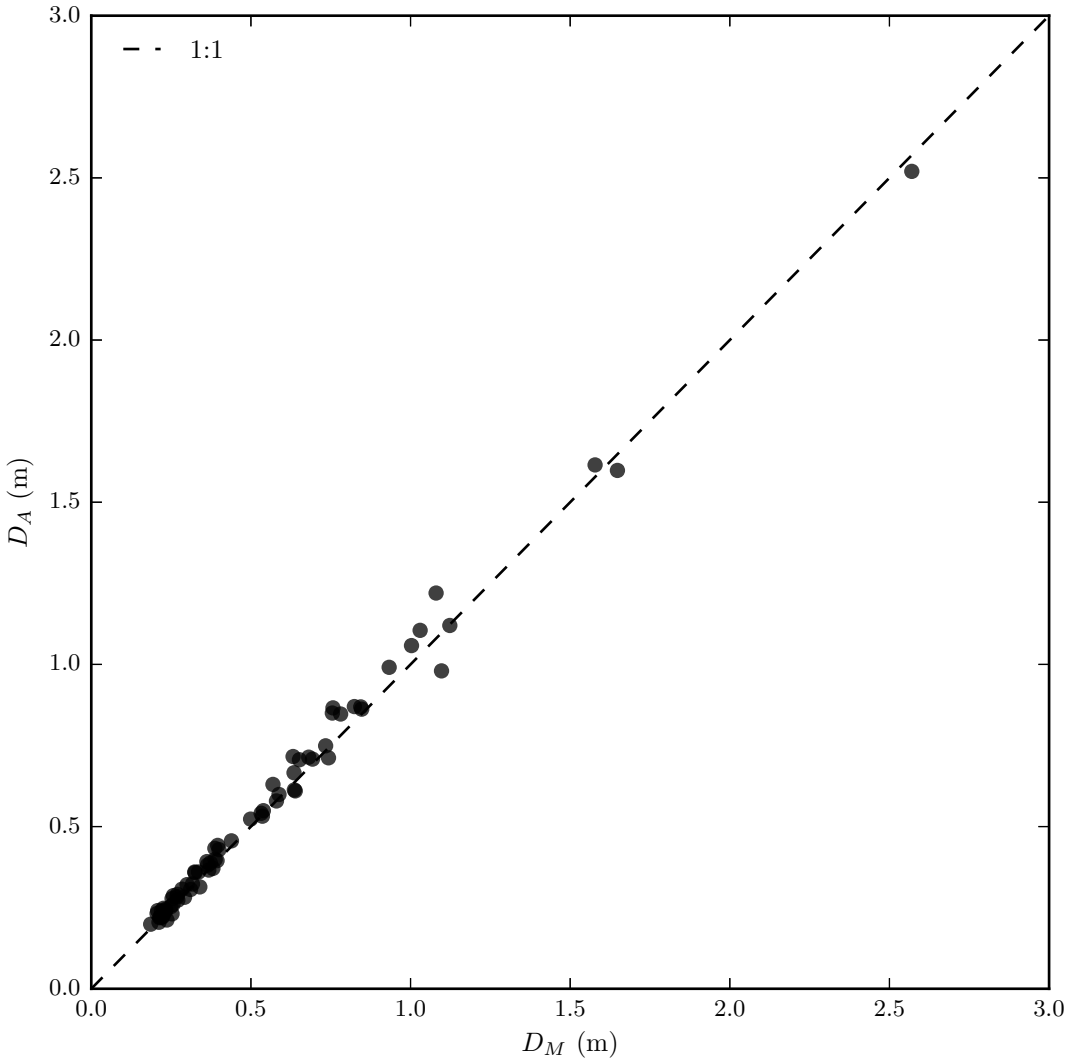


Figure 5.48: Comparison of D retrieved from 75 tree-level point clouds via the automatic iterative RANSAC cylinder fitting routine described in section 5.3.8 (D_A) against the manual measurement of D obtained by averaging the diameter of the major and minor axes across a slice of the stem (D_M).

this readily allowed for identification of any points that had been erroneously segmented out of the crown. In addition to this, there was no indication across any of tree-level point clouds that the adopted sample regimes had not permitted a sampling of the tree tops. As a result, the uncertainty in the measurement of H from the TLS data should be approaching the ranging accuracy of the instrument.

5.5 Discussion

5.5.1 Automation

The results presented in section 5.4 have demonstrated that the methods of *treeseg* have enabled the semi-automatic extraction of 441 individuals from 3.25 ha of predominately tropical forest. From these 441 individuals, 342 have been extracted completely automatically from their respective plot-level point clouds. This represents an overall success rate of 77.6 %.

Across the four forest plots to which *treeseg* has been applied, the individual success rate was 92.9 %, 82.2 %, 78.1 % and 71.0 % for plots KARA-001, LPG-01, CAX-A and NOU-11 respectively. This indicates that the success rate is inversely proportional to the structural complexity of the forest plot. Of the 99 individuals that did require manual refinement, this required $\approx$ 10 minutes per tree, although this did vary significantly on a tree by tree basis.

Figure 5.27a illustrated a typical instance in which manual refinement was required. The need for this step as (so far) always been a result of vegetation interaction. More specifically, this was usually due to interaction between trees of different size classes. The most usual cause of this being the crown of a smaller stem competing to emerge into the canopy; this crown would then interact with the low order branching of a larger stem. There were few instances of interaction between the crowns of dominant canopy trees, this is suggestive of some degree of crown shyness across the tropical forest plots that were considered in this thesis. Alternatively, particularly with respect to plot NOU-11, the presence of a number of lianas contributed to a lower success rate.

When this manual refinement step was required, it was to prune erroneously extracted points from the crown only. That is, the stem extraction was always successful. By placing a constraint on the crown extraction that large clusters could not be children of small clusters, the full extraction of neighbouring crowns was prevented. This ensured that the manual refinement was not too laborious. As a result, the time required for the manual refinement of a tree-level point cloud was significantly less than that required if the extraction was to be achieved entirely through 'by hand extraction'. As described in section 5.3.7, the intention here was to extract each individual from the plot-level point cloud in their entirety. If it was acceptable to sacrifice elements of the high order

branching or leafy material, then the hard thresholds that defined the crown extraction could be tightened. This would likely lead to a significant increase in the success rate.

As described in section 5.1, the aim of *treeseg* was to fill one of the knowledge gaps currently prohibiting the wide-scale application of TLS across tens of 1 ha forest plots. This knowledge gap was the automatic extraction of tree-level point clouds from plot-level point clouds to permit access to the tree-level information content. The methods of the *treeseg* algorithm presented in this chapter have largely filled this knowledge gap. Even in the instances when the extraction is not entirely automatic, the extracted tree-level point clouds are such that the significant majority of the extraction has been completed. Improving on this success rate will be exponentially difficult simply because of the complexity and variety of structure across tropical forests (Hallé et al. 1978). As TLS instruments continue to improve, a reduction in beam spread and an increase in both ranging accuracy and measurement rate will permit collection of more uniform point clouds. As described in section 5.2, this will improve the performance and consistency of the underpinning techniques of *treeseg*, which should improve on the success rate of the extraction.

Aside from automation, *treeseg* was developed with two other underlying design principles in mind. The first of these principles was the aim to be invariant of forest type. It was shown that across the tropical forest plots considered here, the success rate ranged from 71.0–82.2 %. Although this indicates that forest structure does have an influence on the automatic extraction, this range would not be considered unacceptable. When the algorithm was applied to KARA-001, a significantly less structurally complex plot, the success rate exceeds 90 %. This is a good indication that the algorithm will be readily transferable to less structurally complex forest types. Finally, only a single architectural assumption has been made here. That is, trees exit the ground within 35° normal to the local ground vector. This lack of architectural assumptions should help make the algorithm both forest type invariant and non species-specific.

The second design principle was for the algorithm to be instrument agnostic. The methods presented here have used the 3D information of the point cloud (i.e., no reliance on intensity, or other information that it likely to be more instrument-specific). There is no obvious reason as to why these methods would not be transferable to point clouds with a similar ranging accuracy and beam divergence to the instrument used here.

5.5.2 Input parameters

The methods of *treeseg* solely utilise the four techniques of Euclidean clustering, principal component analysis, region growing segmentation and RANSAC cylinder model fitting, as described in section 5.2. It was demonstrated in section 5.2 that the consistent performance of these techniques was linked to variations in nearest neighbour distance across the plot-level point cloud. It was also demonstrated that across plot-level point clouds acquired from forest scenes, particularly tropical forest scenes, the variation in nearest neighbour distance can transcend more than an order of magnitude. This variation is a systematic response to the position in the canopy (i.e., height), such that it is a function of the angular resolution, the beam spread and the general arrangement of the forest elements that leads to the occlusion of other elements. To partially compensate for this variation, the height-resolved mean nearest neighbour distance, described in section 5.2.1, has been used to drive these four techniques.

The variation in nearest neighbour distance was also shown to be a response to the local arrangement of the forest elements such that particular elements in the plot were occluded beyond the plot-level average. To compensate for this local variation and to significantly reduce the variation across the height-resolved mean nearest neighbour distance, downsampling was employed. Probably the most significant input parameter into *treeseg* is the selection of an appropriate value of voxel edge length used in the downsampling. Across all the data considered here, a consistent edge length of 0.04 m was used. Figure 5.15 demonstrated the impact this downsampling had on the plot-level point cloud. This edge length was considered appropriate here as it gave the best trade-off between minimising the variation in nearest neighbour distance without too severely reducing the resolution of the point cloud. It can be seen in figure 5.15 that the downsampling has reduced the range in the height-resolved mean nearest neighbour distance across the lower 30 m of the plot-level point cloud by over 40 %. Implicitly then, this has also significantly curtailed localised variation in the nearest neighbour distance across the point cloud.

Figure 5.15 also superimposed the diameter of the beam footprint as a function of height. The dashed line represents the smallest unambiguous distance in the point cloud regardless of either the angular resolution or degree of occlusion. This figure illustrates

that the vast majority of the data pruned as a result of the downsampling came from redundant points in the hugely oversampled first 20 m. Post-downsampling, the smallest resolvable distance in the point cloud is 0.022 m. With the indented application of tree-scale retrieval of AGB from the extracted tree-level point clouds, this was considered an acceptable loss of resolution in order to permit a consistent performance in the extraction.

The four techniques underpinning *treeseg* described at the beginning of the section compose of 6 input parameters other than the plot-level point cloud. Of these 6 parameters, 3 are automatically determined by the height-resolved mean nearest neighbour distance. The remaining 3 input parameters were arrived at through trial and error. Importantly however, as described in section 5.3, they were held constant across each of the plots. This helps ensure that the algorithm only requires modest adjustment of parameters when ingesting a new plot-level point cloud.

5.5.3 Comparisons with the field data

The comparisons between the TLS data and field data in section 5.4 highlighted a number of inconsistencies between these data. First was the absence of a significant number of stems in the TLS data across the D range between 0.2–0.3 m. In total, there were 137 fewer trees in the TLS data across this bin than were present in the field data. This is probably due to the automatic RANSAC cylinder fitting described in section 5.3.4. As was demonstrated in figure 5.48, there appeared to be a tendency for this approach to underestimate the value of D. With so many of the stems in this range skewed towards the lower value, significant numbers would have pruned as a result of a systematic underestimation in the RANSAC cylinder fit.

Above $D = 0.4$ m there appeared to be a reasonable consistency between the TLS and field data. However, of note was the absence in the TLS data of 2 individuals in CAX-A with $D > 1$ m. A manual check was performed through the slice that was used to identify the stems but neither could be discerned. There was an opposite story to this in LPG-01 where there were 5 individuals with $D > 1$ m present in the TLS data that were absent in the field data. As demonstrated in figure 5.48, there was no evidence that the TLS-derived estimate of D were ever massively overestimated, that would have caused this. One possibility here would be errors in the field data. This has not been explored here, but if there were errors in the field data, then there could be some implication of

this on the aforementioned missing stems in the TLS data across the D range between 0.2–0.3 m.

The comparison of H between the TLS data and field data is particularly interesting in that unlike the distribution of stems by D, there appeared to significant discrepancies. It was seen in both figures 5.40 and 5.47, that the distribution of stems by H across the TLS and field data had a skewed Gaussian appearance. However, in both instances, the mode value of the TLS data was significantly higher than the field data in both forest plots LPG-01 and NOU-11. As there is confidence in the TLS-derived estimates of H, this would indicate a large systematic underestimate in H in the field data. This has significant implications because it was assumed in the study in chapter 2 that all the measurement uncertainties were normally distrusted about zero. As a consequence of this assumption, at the plot-level, the contribution to total allometric uncertainty by the measurement uncertainty was insignificant. However, if the uncertainties in the measurement of height are not normally distributed about zero, then it can be expected for the tree- and plot-scale uncertainties in AGB estimates to be even larger, and more biased, than expected.

Finally, the comparisons of the stem maps between the TLS data and field data were highly inconsistent. Only in LPG-01 and NOU-11 could a fewer of the larger individuals be seen to be in common. This was as anticipated as the definition of x and y in the field data are largely inconsequential.

5.5.4 Further work

There are currently two primary limitations to the *treeseg* algorithm. The first limitation is that in this study, attempts were to made to extract only stems with $D \geq 0.2$ m from the data. In traditional field measurements, and for all the field data considered in this thesis, all stems with $D \geq 0.1$ m are censused. This has been justified here as the interest with the tree-level point clouds lies with tree- and plot-scale retrieval of AGB. From the four sites that were assessed in the experimental study presented in chapter 2 (ANK-01, LPG-01, MNG-03 and MNG-04), the mean contribution of stems in the D range between 0.1–0.2 m was 5.4 % of plot-scale AGB. Therefore, the contribution of the stems in this range to the retrieval of plot-scale AGB is minor.

However, further work is required to extract all trees with $D \geq 0.1$ m to demonstrate

that the methods presented here can be applied all stems across a forest plot that would be traditionally considered. There is no obvious limitation as to why the algorithm could not be successfully applied to smaller stems, it was just not considered in the work for this thesis. Indeed, it may be easier to extract the smaller stems as their crowns are often likely to be at a closer range to the scan locations than the crowns of the larger stems, so this should result in a better definition of these crowns. The second limitation has been discussed in section 5.5.3. That is, the stems absent in the TLS data across the D range from 0.2–0.3 m. This requires further work to fully establish why this is occurring.

One avenue of further work is the advancement of the leaf pruning method described in section 5.3.9. It was demonstrated that the method could perform a rudimentary classification of wood and leaf returns. From this classification, separate wood and leaf-only point clouds were produced for the example tree-level point cloud. This has potentially broad application. For instance, concave hulls could be fitted to the leaf-only point cloud to estimate the volume occupied by the leafy material of the tree; this could in turn be calibrated to provide an estimate to the crucial biophysical parameter, leaf area index. A wood-only point cloud would help reduce the interference of leafy material in woody structure retrieval via geometric primitive fitting. This would have potential to reduce the uncertainties in associated products, such a tree-scale volume and AGB. The caveat to this method was the requirement for the definition of the three hard thresholds used to classify the form of the clustered point clouds. The thresholds that have been used here were identified by trial and error; additionally they required varying on a tree by tree basis. Further work would be required to automatically define these thresholds. A prime candidate to achieve this could be supervised machine learning (Goldberg et al. 1988).

5.6 Conclusions

This chapter has presented *treeseg*, an algorithm that has been developed during the work for this thesis to bridge the first of the knowledge gaps identified in chapter 3: the automatic extraction of tree-level point clouds from plot-level point clouds. The methods presented here have applied Euclidean clustering, principal component analysis, region growing segmentation and RANSAC cylinder fitting to assess the underlying surfaces

of deliberately partitioned sections of plot-level point clouds. From these partitioned sections, tree-level point clouds have been grown through connectivity testing.

The algorithm has been applied to the 3.25 ha of predominately tropical forest described in chapter 4 section 4.2. From these data, 441 individuals have been semi-automatically extracted. Of these individuals, 342 were extracted automatically, representing a success rate of 77.6 %. For the remaining 99 individuals, only limited manual refinement was necessary. As a result, the methods described here have reduced both the time requirement and laborious user input that would have been required to achieve this through 'by hand extraction' by several orders of magnitude. This demonstrates that the knowledge gap has been largely bridged.

The following chapter will describe methods that have been developed during the work for this thesis using these 441 extracted stems to bridge the two remaining knowledge gaps identified in chapter 3.

6 Error in the automatic retrieval of tree- and plot-scale AGB

6.1 Introduction

Chapter 2 presented a novel experimental study demonstrating that relative uncertainties in pan-tropical allometric models forecasting above-ground biomass (AGB) exceeded 80 % and 25 % at the tree- and plot-scale respectively. Due to these uncertainties, it was concluded in this chapter that there is a requirement for the introduction of alternative methods.

Chapter 3 presented a review of terrestrial laser scanning (TLS), the alternative method considered in this thesis for estimating tree- and plot-scale AGB. It was shown in chapter 3 section 3.1 that commercial off-the-shelf instruments are readily capable of the accurate 3D sampling of a forest scene, realised through the construction of a plot-level point cloud. It was also shown in chapter 3 section 3.2 that new modelling techniques fitting geometric primitives to tree-level point clouds are capable of faithfully reconstructing the surface topography of tree-scale woody structure in its entirety (Raumonen et al. 2013; Hackenberg et al. 2015). The resultant quantitative structural models (QSMs) allow estimation of tree-scale volume, that, when accompanied by an estimate of average wood density, permit the retrieval of tree-scale AGB.

Chapter 3 section 3.2 also reviewed the two pilot studies combining these new methods with real and simulated TLS datasets where tree-scale volume or AGB was known *a priori* (Disney et al. 2012; Calders et al. 2015). These two studies demonstrated that tree-scale volume and AGB could be accurately estimated by these new TLS-derived methods to within 10 % of harvested mass. These new methods then, free from the assumptions that burden their allometric counterparts with such large uncertainties, present a potential paradigm shift in the way that forests are measured.

Chapter 3 section 3.3 concluded by stating the knowledge gaps that currently prevent the wide-scale application of these new methods (here meaning application across tens of 1 ha forest plots). This current restriction on employing these new methods on a wide-scale prohibits them from becoming a viable alternative to traditional allometrics;

and as such, from filling the requirement for the introduction of alternative methods identified in chapter 2. The first of these knowledge gaps, the automatic segmentation of tree-level point clouds from plot-level point clouds, was addressed by the development of the *treeseg* algorithm described in chapter 5.

The first of the remaining two knowledge gaps is the automatic reconstruction and optimisation of the QSMs. Here, automation refers to the removal of any requirement of user input in the reconstruction of tree-level point clouds, so as to eliminate any impracticality occurring when up-scaling these new methods across several thousands of tree-level point clouds. Section 6.2 of this chapter presents a routine that has been developed to automatically optimise QSMs that have been constructed via the algorithm of Raumonen et al. (2013), so as to bridge the first of these remaining knowledge gaps.

The second remaining knowledge gap is the quantification of the uncertainties that are attributed to these new TLS-derived estimates of tree- and plot-scale AGB. It is critical to assess these uncertainties to determine how they compare to the uncertainties in the allometric methods, so as to ascertain whether these new TLS-derived methods are indeed a viable alternative to allometrics. These uncertainties are assessed in section 6.3 by applying TLS-derived methods to validation data acquired from 53 individuals which have been weighed by destructive harvest. This permits an explicit quantification of the error in these new TLS-derived estimates of tree- and plot-scale AGB.

6.2 Methods

To permit retrieval of tree-scale AGB from tree-level point clouds, the shape-fitting algorithm of Raumonen et al. (2013)[1] has been considered in this chapter. This algorithm ingests a tree-level point cloud and aims to segment the cloud into distinct sections of stem and branching to which cylinders are fitted to reconstruct the entire woody structure of the tree. An example of a tree-level point cloud and subsequent QSM that was constructed using the *TreeQSM* algorithm, for a *Hymenaea Courbaril* from Nouragues Nature Reserve, French Guiana (NOU-11) is presented in figure 6.1.

The methods of *TreeQSM* are outlined in section 6.2.1 where it is shown that alongside the tree-level point cloud, the algorithm also requires several user-defined input parameters that play a critical role in the goodness of the resultant QSM. It is demon-

[1]Referred to from here on in as *TreeQSM*.

Figure 6.1: Tree-level point cloud (black) and resultant QSM (red) that was constructed using the *TreeQSM* algorithm (Raumonen et al. 2013) for a *Hymenaea Courbaril* (D = 1.12 m, H = 46.5 m) from Nouragues Nature Reserve, French Guiana (NOU-11).

strated in section 6.2.3 that these input parameters vary significantly on a tree by tree cloud basis (i.e., across a plot, the optimal value of these parameters fluctuates) due to variety of factors. Currently, a trial and error approach is used to define these input parameters through visual inspection of the tree-level point cloud and the QSM (i.e., the user generates a number of QSMs across a distribution of the input parameters and makes a judgement as to which QSM appears to most faithfully represent the tree-level point cloud).

This is a time consuming effort that renders application of *TreeQSM* across several thousands of tree-level point clouds impractical. Even across several hundreds of tree-level point clouds, this user-specific judgement will inevitably introduce some element of inconsistency. As a result, it is a fundamental requirement for these input parameters to be automatically selected if these new TLS-derived methods are to become viable alternatives to traditional allometrics. Section 6.2.4 presents a novel theoretical routine that has been developed to automatically optimise these input parameters.

6.2.1 Quantitative structural models

The methods of the *TreeQSM* algorithm can be found in the publication of Raumonen et al. (2013). In summary, the algorithm aims to segment a tree-level point cloud, P, into sections that distinctly represent either the stem or a singular component of the branching (i.e., singular meaning a section of branching constrained at either end by bifurcation). For each of these sections of point cloud, multiple cylinders are fitted to accommodate for tapering of the stem or branch, and to also allow for localised variation in surface topography, in particular, changes in orientation.

The construction of each QSM commences with the selection of a random point p in P (i.e., $p \in P$). From this point, a set of spheres, $\{B\}$, $\{B\} \subset P$, each of constant radius, r, partition the full extent of P, such that individual spheres, B, persist if they comprise of at least one constituent point. From $\{B\}$, a further set is drawn, $\{b\}$, $\{b\} \subset \{B\}$. The objective of this set is for each subset, b, to represent a section of the local underlying surface of the tree (i.e., a small section of the stem or a neighbouring section of off-shooting branch, but preferably not both). $\{b\}$ is derived from $\{B\}$ on the condition that the minimum distance between the centres of any two neighbouring b is greater than the distance, d. Two further conditions constraining the set $\{b\}$ are that all points inside an arbitrarily selected b have a Euclidean distance to the centre of b less than or equal to d, and that the number of points in this b is greater than or equal to N.

b is referred to as the 'cover' and each 'patch' in the cover (i.e., b, $b \in \{b\}$) now acts as the smallest resolvable component of the tree. That is, P has been reduced to $\{b\}$. The cover across a small section of tree-level point cloud acquired from a Norway Spruce is illustrated in figure 6.2, with the constituent points of each patch in the cover uniformly coloured (Raumonen et al. 2013).

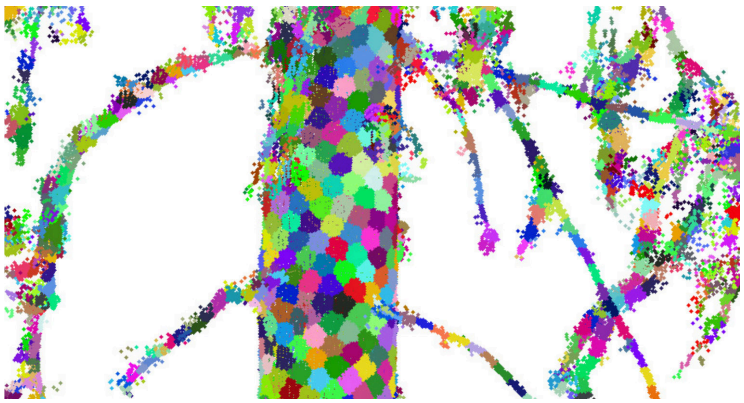


Figure 6.2: The cover generated across a small section of tree-level point cloud acquired from a Norway Spruce, with the constituent points of each patch uniformly coloured (Source: Raumonen et al. 2013).

The next stage of the algorithm is to discriminate each of the patches in the cover based on an assessment of the underlying surface of each patch with respect to the relationships to other neighbouring patches. The underlying surface is inferred through an analysis of the arrangement of the points in each patch. This is achieved via principal component analysis and surface normal estimation, as described in chapter 5 section 5.2.4 and chapter 5 section 5.2.5 respectively. Here, either the primary Eigenvector of the patch, or if that is ambiguous, the surface normal, is used to infer the directionality of the underlying section of stem or branching. If neighbouring patches have a similar directionality then they are considered to belong to the same section of stem or branching. If the directionality significantly diverges from neighbouring patches, then this is used as an indicator of a position of bifurcation.

This segmentation process is achieved by commencing from the base of the stem and noting all the positions of bifurcation. The process is then iteratively repeated from these locations to segment the first, second, third, etc. orders of branching until exhaustion of all the patches in the cover. This segmentation process is demonstrated in figure 6.3 on two small sections of a tree-level point cloud acquired from a Norway Spruce (Raumonen et al. 2013).

The final stage of the *TreeQSM* algorithm is to fit a series of cylinders to each of the stem and branch components that were segmented in the previous step. This is achieved through the non-linear total least squares cylinder fit (Lukács et al. 1998). During this cylinder fitting, several regulations are imposed to prevent the erroneous placement of cylinders, or the incorrect sizing of cylinders, that would arise from either noisy data or

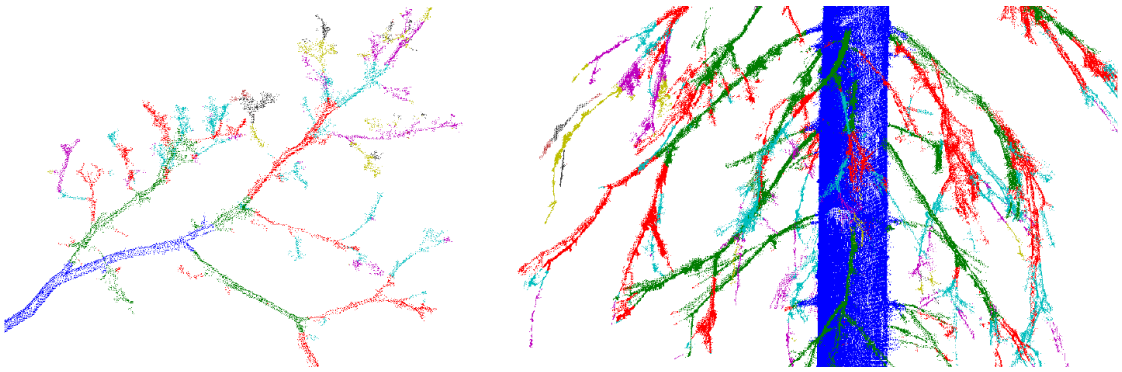


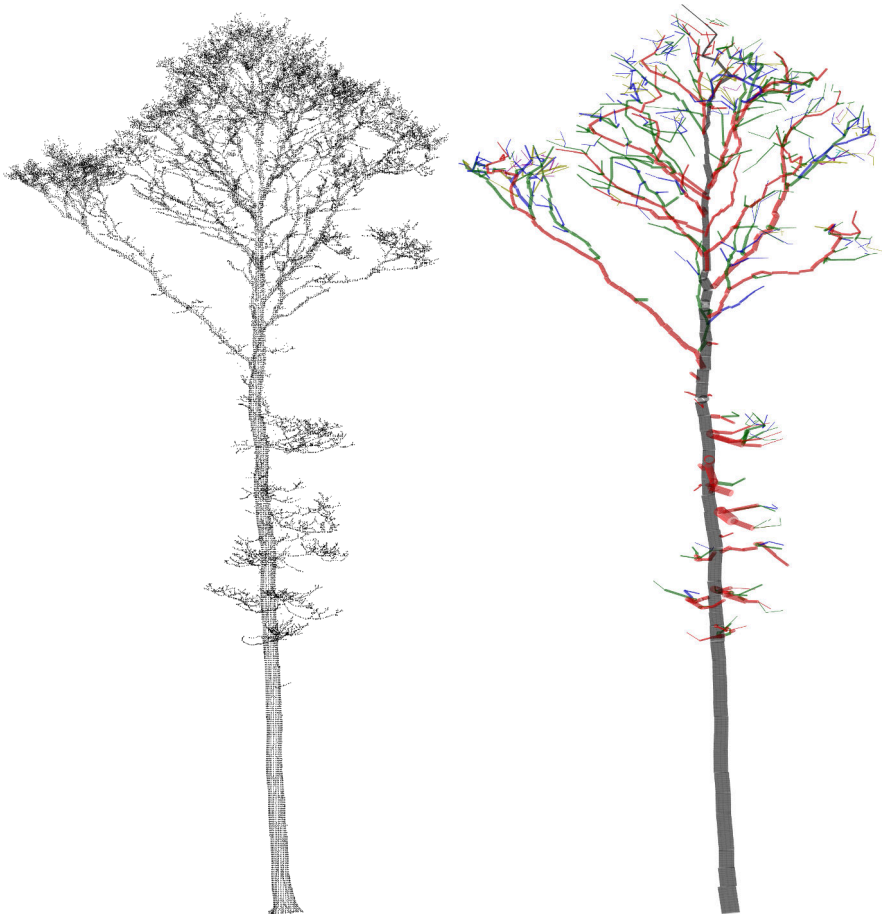
Figure 6.3: Segmentation of the cover, $\{b\}$, via principal component analysis and surface normal estimation, isolating the stem and each order of branching for two sections of tree-level point cloud acquired from a Norway Spruce (Source: Raumonen et al. 2013).

from uncertain inferences that have been drawn from the patches about the underlying surface. For instance, child cylinders will always be smaller than or equal to their parent cylinders.

Figure 6.4a presents a tree-level point cloud representing a *Quercus* spp. individual that was acquired during the work for this thesis from Alice Holt Forest, UK (AH-01). Figure 6.4b then presents the QSM that has been constructed from this tree-level point cloud using the *TreeQSM* algorithm, where each order of branching, inferred during the segmentation stage, has been uniquely coloured.

As can be seen, the algorithm enables the reconstruction of tree-level woody structure all the way through to high order branching. In this instance, it appears that *TreeQSM* is capable of faithfully reconstructing the woody structure of the tree through to at least the third-order of branching.

Alongside this design requirement for *TreeQSM* to permit the entire reconstruction of woody tree structure, it was also designed to be both automatic and fast. Once the algorithm has been provided with the tree-level point cloud and the user-defined input parameters (i.e., d and N), no further user input is required during the reconstruction process. In the case of the QSM presented in figure 6.4b, the QSM was constructed on a sub-minute time-scale using a standard workstation (24-core Intel Xeon E7 1.2 GHz, 120 GB RAM). In addition to this, the algorithm makes very few assumptions about tree architecture, making it broadly applicable across species and biome. Finally, only the 3D information of the tree-level point cloud is considered (i.e., properties such as intensity, which may be available depending on instrument); this permits generic application of the algorithm across TLS data collected from a variety of instruments.



(a) (b)

Figure 6.4: (a) Tree-level point cloud acquired from a *Quercus* spp. individual from Alice Holt Forest, UK (AH-01); (b) the resultant QSM constructed from this tree-level point cloud using the *TreeQSM* algorithm, coloured by branching order.

6.2.2 QSM goodness

Together with the tree-level point cloud, *TreeQSM* requires two primary user-defined input parameters, namely d and N. It is the automatic selection of these two parameters that knowledge gap ii), described in section 6.1, is concerned with.

In this chapter, the interest in the application of the QSMs lies with the estimation of tree-scale volume for the retrieval of tree-scale AGB. As such, the interest in the influence exerted by the input parameters d and N on the goodness of the QSMs is with respect to their impact on the accuracy to which tree-scale volume can be estimated. That is, for instance, rather than the accuracy to which high order branching can be faithfully reconstructed, as would perhaps be more important if the QSMs were to be applied in radiative transfer simulations. In this section, only the accuracy to which tree-scale volume can be retrieved from *TreeQSM* is considered in this goodness assessment. That

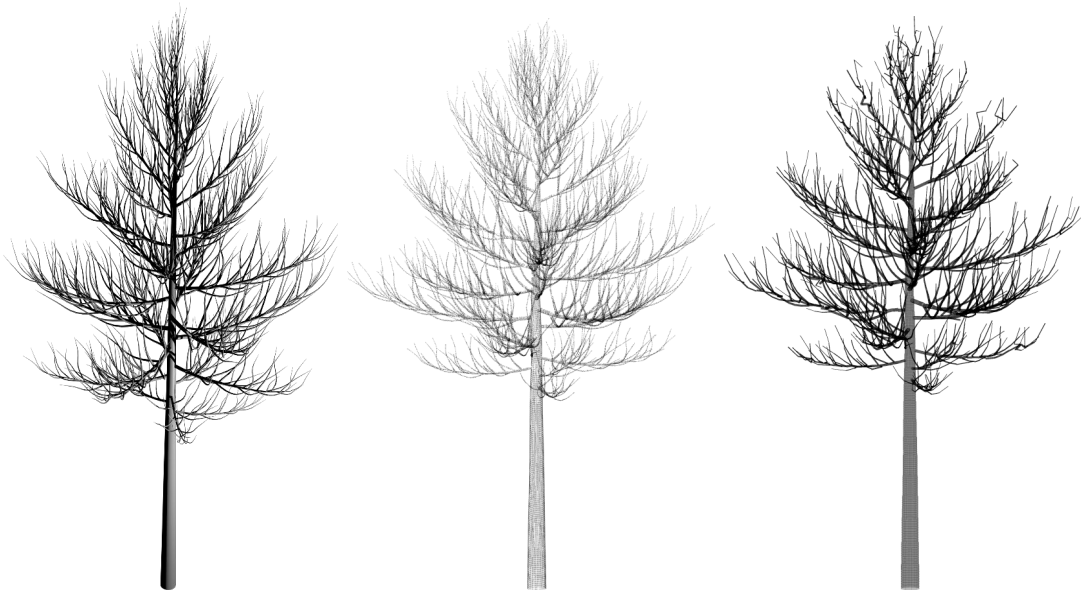
is, the uncertainties associated with subsequently derived estimates of tree-scale AGB, such as the appropriateness of cylinders as the geometric primitive of choice, buttress formations, heart rot and variable woody density are considered in chapter 7.

As described in the previous section, the two user-defined input parameters d and N drive the generation of the cover. That is, d dictates the size of the spheres and N determines whether each sphere contains sufficient points to constitute a patch. By definition then, the selection of these parameters is wholly dictated by the composition and arrangement of the input tree-level point cloud. For example, across a tree-level point cloud, P, with a uniform nearest neighbour distance (see chapter 5 section 5.2.1) of d_{NN}, then when $d = d_{NN}$, each point, p, $p \in P$, will be contained in its own unique sphere when $N = 1$. As the segmentation process aims to discriminate patches by inferring the state of the underlying surface of the patch relative to neighbouring patches, it would be expected that when $d = d_{NN}$, there would be little viable information gleaned regarding the underlying surface. This is because principal component analysis and surface normal estimation typically require a minimum of several points to return indicative information with respect to the underlying surface. In this example then, d will need to be somewhat larger than d_{NN}, although not too large that local structure, such as bifurcation points are masked by being incorporated into the same patch.

To demonstrate the influence that the user-defined input parameters d and N have over the quality of the QSMs, a perfectly-sampled tree-level point cloud with a uniform d_{NN} has been generated. Such a tree-level point cloud is useful in conducting a review of the impact that d and N have on the QSMs as it permits the decoupling of the potential influences of point cloud data quality on the reconstruction process. To achieve this, a point cloud has been simulated using the tree model illustrated in figure 6.5a. This tree model, generated using the *Sapling Tree* extension (Hale 2011) to *Blender* (Foundation 2017), with a total volume of $1.720\,\mathrm{m}^3$, represents a generic Birch tree with two orders of branching in leaf-off conditions.

TLS measurements have then been simulated to generate a point cloud from this tree model via *librat*, a Monte Carlo ray tracing radiative transfer model (Lewis 1999). In this instance, TLS data acquisition was simulated using four scan positions at a distance of 10 m from the model on the principal axes. At each of these positions, pulses were fired across the full hemisphere, with each pulse tested as to whether it intersected with

the tree model. If an interaction did occur, then the location of this interception was considered a return and added to the simulated point cloud. For this simulation, the characteristics of the RIEGL VZ-400 were mimicked (see chapter 4 section 4.1), such that the pulse has beam divergence of 0.35 mrad and an angular sampling pattern of 0.04°. Using this simulated sampling regime, the tree model was near-perfectly sampled. To recreate the scenario described above, where a uniform d_{NN} is present, the simulated tree-level point cloud was then downsampled via voxel grid downsampling (see chapter 5 section 5.2.3) using a voxel edge length of 0.04 m. The resultant tree-level point cloud is shown in figure 6.5b.



(a) **(b)** **(c)**

Figure 6.5: (a) Procedurally generated tree model created using the *Sapling Tree* extension to *Blender*, of a generic Birch tree with two orders of branching in leaf-off conditions; (b) the simulated point cloud generated from the model of (a) using the *librat* ray tracer and subsequently downsampled via voxel grid downsampling; (c) the QSM constructed from the simulated point cloud in (b) using the *TreeQSM* algorithm.

Using this simulated tree-level point cloud, a number of QSMs were created using a range of values of d from 0.03–0.1 m, with a step of 0.01 m, whilst N was held constant at 3. Figure 6.6 illustrates three of these QSMs constructed using values of d of 0.03 m, 0.06 m and 0.1 m respectively. In figure 6.6a, where $d = 0.03$ m, it can been seen that the stem has not been well reconstructed, with a number of smaller cylinders, of seemingly random sizing and orientation, being the representatives of the stem. In addition, it can be seen that the reconstruction of the woody structure of the crown appears somewhat underdeveloped. In figure 6.6b, where $d = 0.06$ m (i.e., d is 50 % larger than the

downsampling voxel edge length), it appears that the QSM has faithfully reconstructed the woody structure of both the stem and the crown. In figure 6.6c, where $d = 0.1\,\text{m}$, the stem also appears to have been consistently modelled, however the crown now appears overdeveloped with many branches larger than their counterparts observed when $d = 0.06\,\text{m}$.

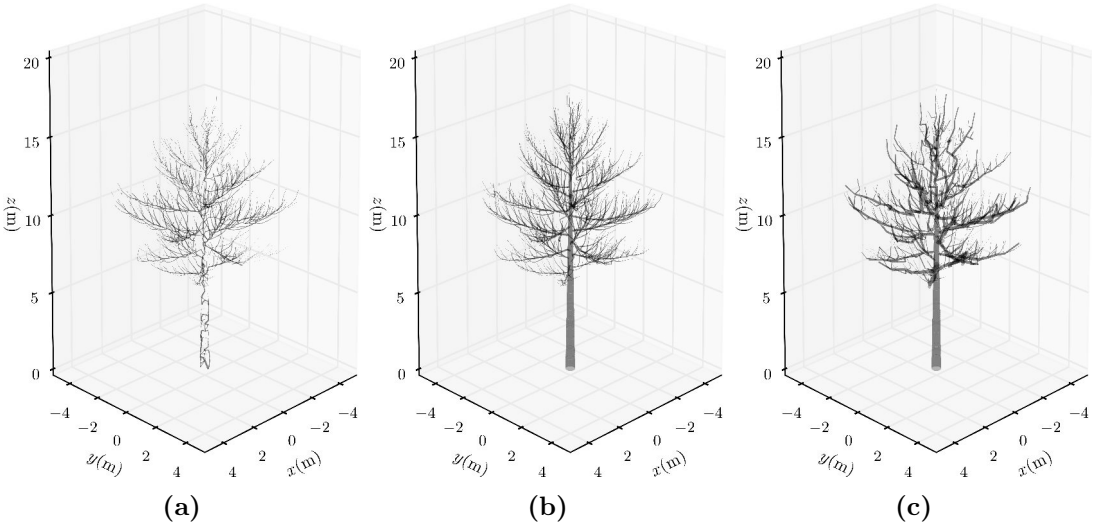


Figure 6.6: QSMs generated from the tree-level point cloud shown in figure 6.5b, where (a) $d = 0.03\,\text{m}$, (b) $d = 0.06\,\text{m}$ and (c) $d = 0.1\,\text{m}$; N was held constant at 3.

The distribution of estimated tree-scale volume returned from the QSMs generated across the considered range of values of d from 0.03–0.1 m is presented in figure 6.8. As was described in section 6.2.1, the starting location from which the cover generation commences is random. Thus, for a constant tree-level point cloud and a constant set of user-defined input parameters, there will be a variety of permutations of the cover. For each of these possible permutations then, each cover will produce a different segmentation, that will, in turn, produce a different set of fitted cylinders. This results in variable estimates of tree-scale volume for each iteration of *TreeQSM*, despite the tree-level point cloud and input parameters being held constant.

To demonstrate the influence of this on the estimates of tree-scale volume, figure 6.8 also presents, for each value of d, the 95 % confidence intervals obtained from the construction of 20 QSMs. That is, the mean volume, $\bar{V}$, from these 20 QSMs is defined as:

$$\bar{V} = \frac{1}{N} \sum_{i=1}^{N} V_i \tag{6.1}$$

With the standard deviation of the volume, σ_V, across the 20 QSMs defined as:

$$\sigma_V = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(V_i - \bar{V})^2} \tag{6.2}$$

Where the 95 % confidence intervals (i.e., the 95 % probability that V of any further constructed QSM will lie inside the quoted bands), have been constructed using the standard score, $z_{0.025}$, as:

$$\bar{V} \pm \sigma_V \, z_{0.025} \tag{6.3}$$

20 QSMs have been constructed for each value of d as this captures the majority of the observed variation in V. This is demonstrated in figure 6.7, where 50 QSMs have been constructed from the simulated tree-level point cloud shown in figure 6.5b, where $d = 0.06$ m and $N = 3$. Shown in the figure, is σ_V that has been consecutively calculated, from a random selection of 3–50 of these QSMs; it can be seen that σ_V largely stabilises when it is calculated from 15 or more QSMs.

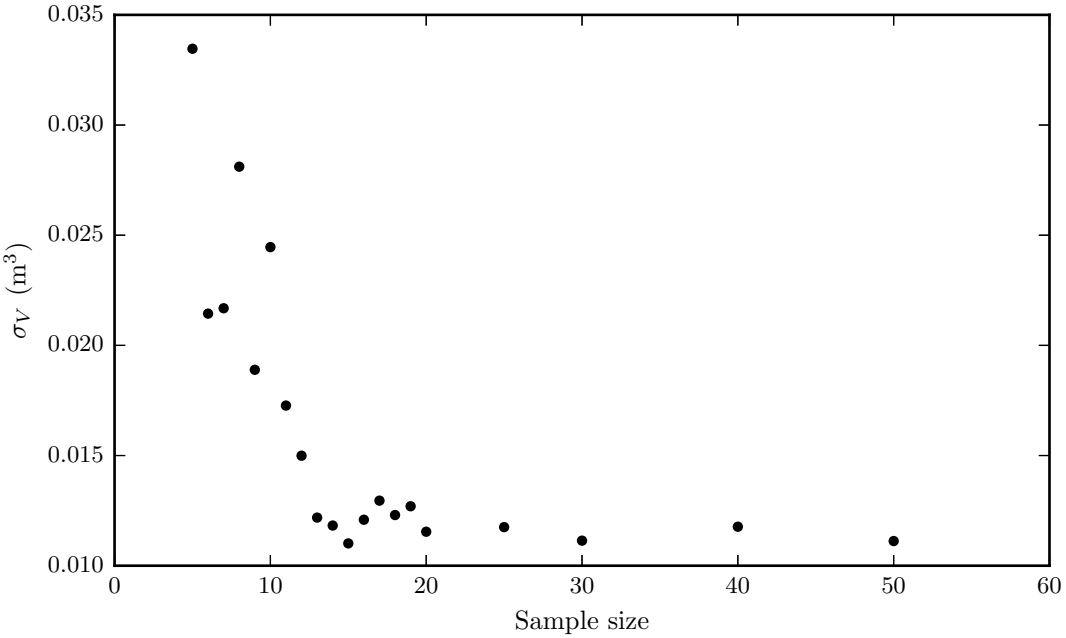


Figure 6.7: The standard deviation of the volume, calculated from the construction of 50 QSMs from the simulated tree-level point cloud shown in figure 6.5b, where $d = 0.06$ m and $N = 3$; σ_V has been calculated for a random selection of 3–50 of these QSMs.

It can be seen in figure 6.8 that the user-defined selection of d can have two profound

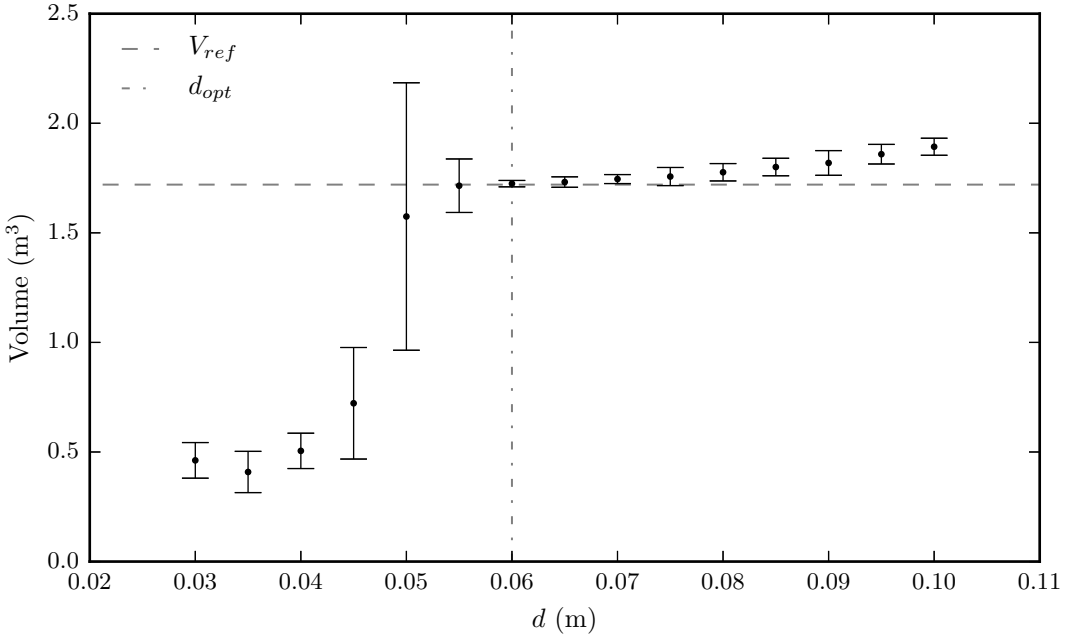


Figure 6.8: Mean estimates of tree-scale volume derived from 20 QSMs constructed from the simulated tree-level point cloud illustrated in figure 6.5b per value of d across the range 0.03–0.1 m, where $N = 3$; the intervals represent the 95 % confidence intervals.

impacts on the QSM-derived estimates of tree-scale volume. First, across the considered range of values of d, tree-scale volume ranged from 0.408–1.892 m^3. The value of $d =$ 0.06 m produced the estimate of volume closest to the actual volume, coming in at 1.725 m^3, representing a slight overestimate with a percentage error of 0.3 %.

The second and perhaps more important impact of d on the QSM-derived estimates of tree-scale volume is the variability in the size of the confidence intervals. The relative size of the confidence intervals, r_{CI}, is defined as:

$$r_{CI} = \frac{\sigma_V \, z_{0.025}}{\bar{V}} \tag{6.4}$$

In this instance, r_{CI} ranged from 2.5–38.0 %. The minimum r_{CI} occurred when $d =$ 0.06 m, coinciding with the value of d that provided the best estimate of tree-scale volume. It can be seen that on either side of this value of d, r_{CI} begins to increase, more so for the lower values of d.

This behaviour is quite intuitive when considered with respect to the generation of the cover and the possible permutations that the cover can take. For example, in an extreme scenario, such that d is so large that the cover comprises of just 50 patches, then

the arrangement of these patches will be markedly different depending on the randomly selected starting location of the cover generation. This also holds for small values of d, where each permutation of the cover can lead to a drastically different arrangement of the patches. These potentially different arrangements of the cover will lead to variable segmentations and subsequent cylinder fits, that will result in variable estimates of the tree-scale volume. As d approaches a value where the cover relatively well defines the underlying local surfaces of the tree, and the locations of bifurcation, then the possible permutations of the cover become more uniform. Intrinsically then, it would be expected that r_{CI} would begin to narrow as d approaches the optimal value, as demonstrated in figure 6.8.

The second user-defined input parameter considered here is N; the minimum number of points inside a sphere to be considered a patch, and thus a constituent of the cover. As with the consideration of the input parameter d, the simulated tree-level point cloud illustrated in figure 6.5b was reconstructed across a range of values of N from 1–15 , whilst d was held constant at 0.06 m. Figure 6.9 presents three of these QSMs constructed using values of N of 1, 3 and 15 respectively.

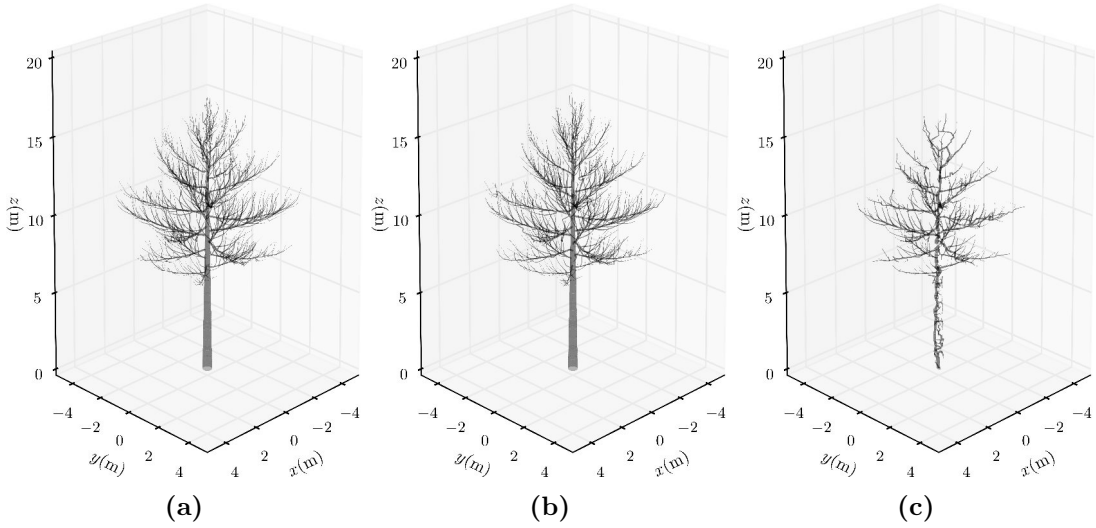


Figure 6.9: QSMs generated from the simulated tree-level point cloud shown in figure 6.5b, where (a) $N = 1$, (b) $N = 3$ and (c) $N = 15$; across the three reconstructions, d was held constant at 0.06 m.

It can be seen from this figure that alongside d, the appropriate selection of N is also influential in the quality of the QSMs. For instance, in figure 6.9c where $N = 15$, both the stem and crown have been incorrectly reconstructed, in a similar fashion as to when a small value of d was selected. In figures 6.9a and 6.9b, where values of N of 1

and 3 were used respectively, the QSMs are almost indistinguishable. It can just about be seen that when $N = 1$, at the top of the crown, the orientation and linkage between cylinders appears slightly less consistent than when $N = 3$.

Figure 6.10 presents the distribution of estimated tree-scale volume returned from the QSMs generated across the considered range of values of N from 1–15 Included in the figure are the 95 % confidence intervals derived from the 20 QSMs constructed per value of N.

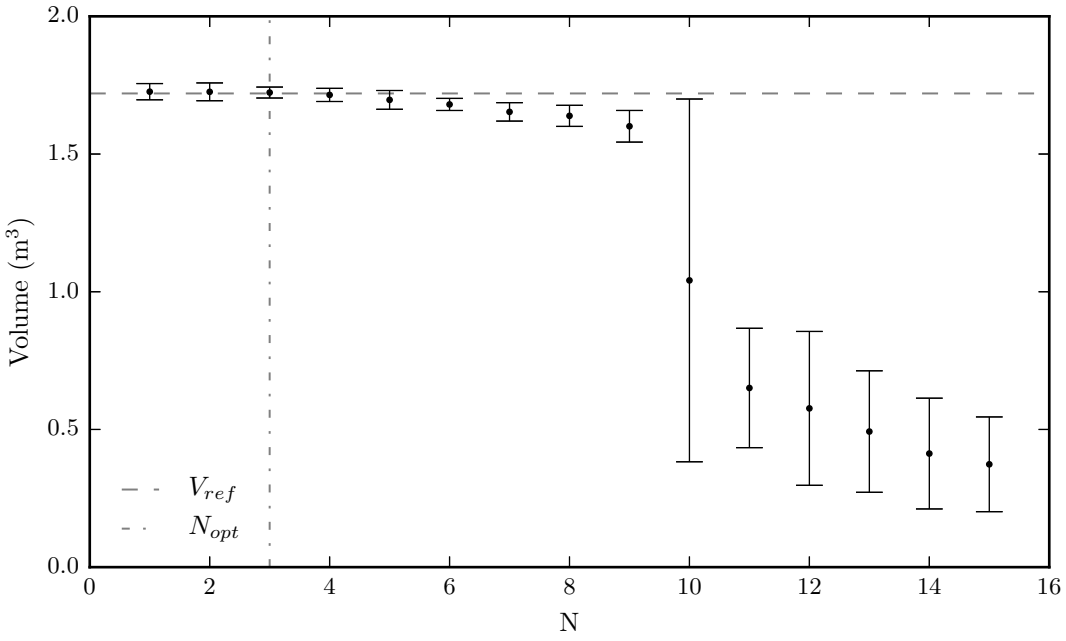


Figure 6.10: Mean estimates of tree-scale volume derived from 20 QSMs constructed from the simulated tree-level point cloud illustrated in figure 6.5b per value of N across the range 1–15 , where $d = 0.06$ m; the intervals represent the 95 % confidence intervals.

Figure 6.10 demonstrates that N also influences both the returned tree-level volume and r_{CI}. Across the range of values of N from 1–3 , it can be seen that volume remains largely consistent at approximately 1.725 m^3. Above these values, the volume slowly trends to lower values prior to a significant decrease in volume when $N = 10$. r_{CI} ranged from 2.5–56.0 %, with the minimum occurring when $N = 3$. It can be seen that either side of this minimum, r_{CI} sequentially increases.

Similar to the behaviour of d, the response of tree-scale volume to variation in the user-defined input parameter N is also intuitive. In this example, d was held constant at 0.06 m, therefore, with this perfectly-sampled tree-level point cloud with uniform d_{NN}, this value of d will typically be 50 % larger than d_{NN} across the point cloud. As such, if

the condition of $N > 6$ is imposed (i.e., the cubic neighbours), then it would be expected that the cover will comprise fewer patches and not be a complete representation of the full extent of the point cloud. This will, in turn, result in the fitting of fewer cylinders and manifest as a lower returned tree-scale volume. Likewise, when $N = 1$, it can be presumed that at the tips or edges of branching, that a number of the patches will persist that contain only a single point. As such, inferences drawn from such patches regarding the underlying surface will potentially be more ambiguous. This explains the less consistent orientation and linkage of the cylinders representing the higher order branching.

Figures 6.8 and 6.10 have demonstrated that the appropriate selection of the user-defined input parameters, d and N, is a fundamental requirement for the accurate derivation of tree-scale volume from QSMs derived via the *TreeQSM* algorithm. The remaining factor influencing the accurate retrieval of tree-scale volume from the QSMs that requires consideration is the tree-level point cloud itself. Up to this point, only the simulated tree-level point cloud illustrated in figure 6.5b, generated through a near-perfect simulated sampling regime and downsampled to achieve a uniform d_{NN}, has been considered.

The quality of a plot-level point cloud was defined in chapter 3 section 3.1.5 as the completeness to which the point cloud captures the 3D structure of the scene, within the bounds of fixed instrument-specific parameters. To a tree-level point cloud, this is the completeness to which the 3D woody and leafy structure of the tree in question has been captured.

As was demonstrated in chapter 5 section 5.2.1, in structurally complex forest scenes such as tropical forests, the quality of tree-level point clouds significantly diminish when ascending from the forest floor through to the top of the canopy. This was demonstrated in chapter 5 section 5.2.1 figure 5.3 by the height-resolved mean nearest neighbour distance, $\overline{d_{NN}}(z)$, through the plot-level point cloud acquired from plot NOU-11. This variation was shown to be a result of increasing range between the scanner and the target, manifesting in a larger step distance between sequentially fired pulses and a larger beam footprint. It was also shown to be due to the structural arrangement of the forest, with general occlusion of canopy elements by other elements in the undergrowth and understory. It was demonstrated in chapter 5 section 5.2.3 that point clouds acquired in tropical forests also regularly suffer from localised variation in d_{NN}, far beyond that

of $\overline{d_{NN}}(z)$, which is a direct result of significant localised occlusion (e.g., the restricted view of a section of a particular stem due to surrounding lianas). These two factors culminate in tree-level point clouds acquired from tropical forests being an incomplete sample of the tree, whilst also containing a non-uniform distribution of points through the extent of the point cloud.

To demonstrate how the *TreeQSM* algorithm and the resultant QSMs respond to such tree-level point clouds acquired from complex forest scenes, the full tree-level point cloud of a dominant canopy tree, illustrated in Figure 6.11a, is considered here. This tree-level point cloud represents a $D = 1.12\,\text{m}$, $H = 46.5\,\text{m}$, *Hymenaea Courbaril* individual from plot NOU-11. Figure 6.12 presents $\bar{d_{NN}}(z)$ across this point cloud, where it can be seen that $\overline{d_{NN}}(z)$ varies over more than an order of magnitude from 0.003–0.0558 m. Figure 6.13 then presents $\bar{V}$ and 95 % confidence intervals, derived from 20 QSMs constructed for each value of d ranging from 0.03–0.1 m, where $N = 3$. It can be seen that across this distribution of values of d, the volume ranges from 9.722–50.761 m^3, whilst r_{CI} ranges from 21.9–119.9 %. The minimum value of r_{CI} occurs at $d = 0.08\,\text{m}$, resulting in an estimate of tree-scale volume of 46.478 m^3. It can be seen then, that when transitioning from simulated to real-world data, there has been a large jump in the range of r_{CI}.

Concurrently, as shown in figure 6.11b, this same tree-level point cloud has been downsampled via voxel grid downsampling using a voxel edge length of 0.04 m (this is the same voxel edge length that was employed during the downsampling stage of *treeseg*, as justified in chapter 5 section 5.5.2). Alongside the distribution of $\overline{d_{NN}}(z)$ across the original tree-level point cloud, figure 6.12 also presents $\overline{d_{NN}}(z)$ across the downsampled tree-level point cloud. It can be seen that across the downsampled point cloud, $\overline{d_{NN}}(z)$ ranges from 0.022–0.059 m, a decrease in the absolute range of $\overline{d_{NN}}(z)$ by 37.9 % over the original point cloud. It can also be seen that as a result of the downsampling, in contrast to the original point cloud, across the first 27 m in height from the forest floor, $\overline{d_{NN}}(z)$ remains almost constant.

Overlaying $\bar{V}$ and the 95 % confidence intervals of the original point cloud in figure 6.13, are the counterpart values for the downsampled point cloud. Similar to the original point cloud, these mean volumes and 95 % confidence intervals have been derived from 20 QSMs constructed for each value of d ranging from 0.04–0.15 m, where $N = 3$. However, unlike the original point cloud, those cylinders with a radius below 0.025 m have been

pruned from the QSMs constructed on the downsampled tree-level point cloud. This is because, as illustrated in figure 6.12, the smallest d_{NN} as a result of the downsampling is now 0.022 m. As such, any cylinders sized below this minimum distance will have been constructed upon ambiguous inferences and as such, require discarding.

It can be seen in figure 6.13, that the distribution of $\bar{V}$ for the downsampled point cloud ranges from 0.779–61.669 m^3, whilst r_{CI} ranges from 7.4–109.5 %. The smallest value of r_{CI} occurred at $d = 0.1$ m, resulting in an estimate of tree-scale volume of 46.055 m^3.

(a) **(b)**

Figure 6.11: (a) The tree-level point cloud representing a $D = 1.12$ m, $H = 46.5$ m, *Hymenaea Courbaril* dominant canopy tree inside plot NOU-11; (b) the same tree-level point cloud post-downsampling, achieved via voxel grid downsampling using a voxel grid edge length of 0.04 m.

It can be seen from figure 6.13 that downsampling the tree-level point cloud to make $\overline{d_{NN}}(z)$ as uniform as possible has significantly reduced the spread of r_{CI}, and consequentially, improved the confidence in the derived estimate of tree-scale volume. In comparison with the original tree-level point cloud, at the minimum observed values of r_{CI}, the intervals have shrunk from 21.9 % to 7.4 % as a result of the downsampling. Again, the underlying reasoning for this reduction in r_{CI} is intuitive. By definition, the more uniform the distribution of points inside the tree-level point cloud, the more likely it is that any set of arbitrarily selected patches will contain a more consistent number of points inside each patch. As such, there will be fewer possible permutations of the

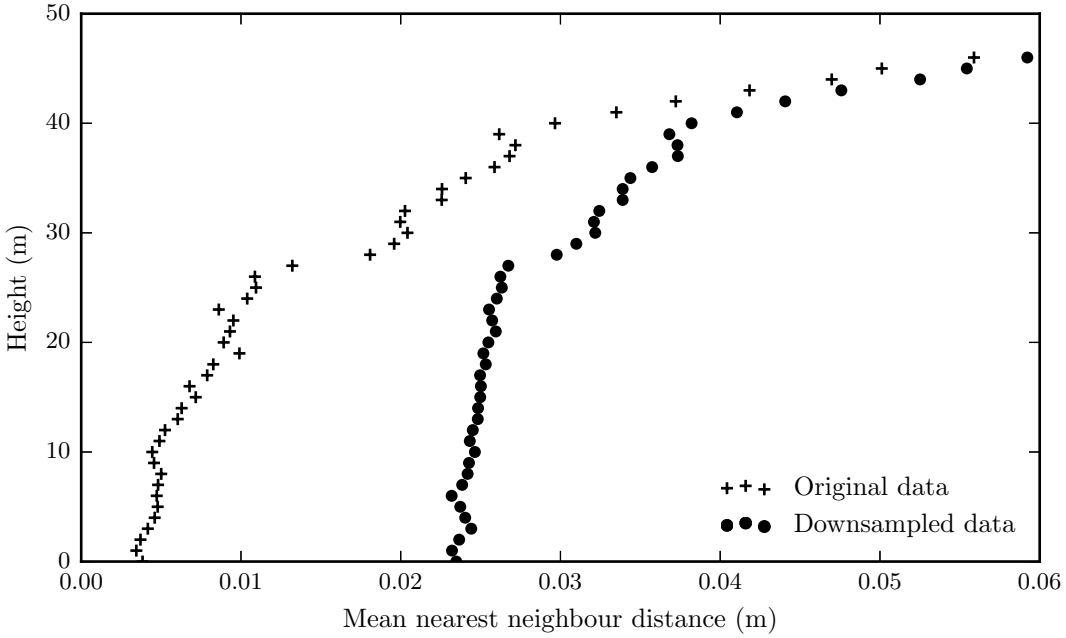


Figure 6.12: Height-resolved mean nearest neighbour distance, $\overline{d_{NN}}(z)$, across the original and downsampled tree-level point clouds shown in figures 6.11a and 6.11b respectively.

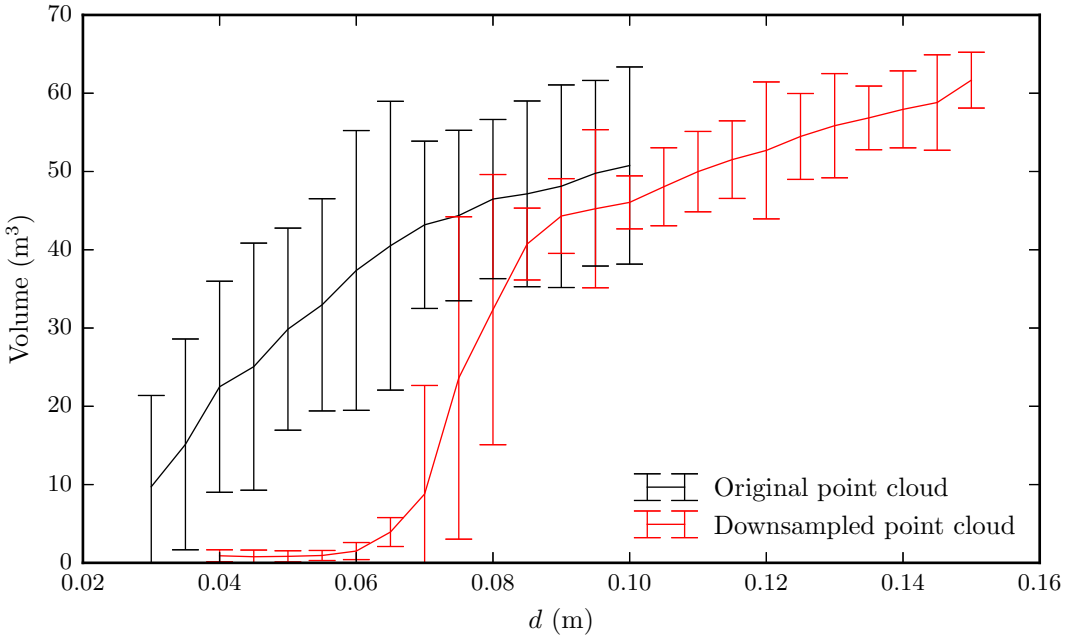


Figure 6.13: Mean estimates of tree-scale volume generated from the original (black) and downsampled (red) tree-level point clouds illustrated in figures 6.11a and 6.11b respectively; 20 QSMs have been constructed per value of d across the ranges of 0.03–0.1 m for the original point cloud and 0.04–0.15 m for the downsampled point cloud, where $N = 3$; the intervals represent the 95 % confidence intervals.

cover, leading to fewer and less varied outcomes of the segmentation, the cylinder fits and the subsequent estimates of tree-scale volume. However, as observed in figure 6.12, the

uniformity of $\overline{d_{NN}}(z)$ across both the original and downsampled tree-level point clouds begins to lose any consistency above the height of approximately 27 m. It can therefore be expected that as a result of these fluctuations, the reconstruction of the upper section of crown will remain uncertain across QSMs constructed using either the original or downsampled tree-level point cloud.

Figures 6.14a and 6.14b present two QSMs that have been constructed from the original and downsampled tree-level point clouds shown in figures 6.11a and 6.11b respectively. In both instances, the QSM shown belongs to the parameter set with the lowest value of r_{CI}, whose volume is closest to $\bar{V}$. Across both QSMs, it can be seen that the cylinders carrying the significant proportion of the tree-scale volume appear largely consistent. This, combined with figure 6.13 illustrating that both estimates of tree-scale volume attributed to the lowest value of r_{CI}, fall within the 95 % confidence intervals of each other (46.478 m^3 vs. 46.055 m^3), demonstrates that the loss of resolution associated with downsampling has little influence on the retrieved estimate of tree-scale volume. Indeed, with respect to tree-scale volume, the significant reduction in r_{CI} that is associated with the downsampling far outweighs any loss of resolution. A beneficial consequence of this downsampling and subsequent pruning of the smaller sized cylinders, is the observable reduction in the interference of leaf returns on the reconstruction process alongside a substantial reduction in processing time. This is assessed further in section 6.3.2.

(a) **(b)**

Figure 6.14: Two QSMs constructed from (a) the original tree-level point cloud in figure 6.11a, and (b) the downsampled tree-level point cloud in figure 6.11b; in both instances, the QSM shown belongs to the parameter set attributed with the lowest value of r_{CI}, whose volume is closest to $\bar{V}$; in (b) all cylinders with a radius below 0.025 m have been pruned, as dictated by $\overline{d_{NN}}(z)$ through the tree-level point cloud.

6.2.3 A routine for automatic QSM optimisation

The previous section demonstrated in figure 6.8 that the selection of the user-defined input parameter d is a critical step in the accurate retrieval of tree-scale volume from QSMs constructed via the *TreeQSM* algorithm. It was also demonstrated in the previous section that d_{opt} varies significantly on a tree-level point cloud basis (i.e., intra-plot). The primary cause of this is the variable quality encountered across tree-level point clouds (i.e., the completeness to which the point cloud captures the 3D structure of the tree alongside the distribution of the points inside the cloud), something that is particularly pronounced for those tree-level point clouds acquired from complex forest scenes, such as tropical forests. To a lesser degree, the user-defined selection of N was also shown in figure 6.10 to be somewhat influential on the estimates of tree-scale volume, although it was demonstrated that these estimates did remain broadly consistent when N was held constant. Rather, the selection of N played a more nuanced role in the faithful reconstruction of the higher-order branching.

Currently, end-users of *TreeQSM* make a judgement as to the appropriate value of

d_{opt} based on a visual inspection between the tree-level point cloud and a number of QSMs that have been generated across a range of values of d. The judgement made by the user is which QSM appears to have most faithfully reconstructed the woody structure of the tree. Across several thousands of tree-level point clouds this approach is impractical, whilst the reliance upon user-specific judgement will also likely induce a bias or some inconstancy. Thus, this knowledge gap, the automatic optimisation of QSMs, is a current limitation prohibiting the wide-scale application of these new TLS-derived methods.

This section presents a routine that has been developed during the work for this thesis to automatically select d_{opt}, and to bridge this knowledge gap. In keeping with the design philosophies of both *treeseg* and *TreeQSM*, this routine has been designed to be generic, so as to make it as broadly relevant as possible. That is, the approach is free from assumptions that would make it either species-specific or instrument-specific.

This routine has been developed theoretically and is based on exploiting the expected behaviour of *TreeQSM* in response to the ingested tree-level point cloud. In particular, this routine aims to automatically prescribe the value of d to *TreeQSM*, as determined from an assessment of the tree-level point cloud, that will produce the best permutation of the cover; so as to permit the most accurate segmentation and cylinder fit as possible. Critically, at the heart of this routine are two self-selecting input parameters based on theoretical minima and maxima. Combining this routine with *TreeQSM* then, enables the reconstruction of the entire woody structure of tree-level point clouds without any end-user input.

This theoretical approach is demonstrated in this section to accurately determine d_{opt} for simulated tree-level point clouds, such as that shown in figure 6.5b, where tree-scale volume, and hence d_{opt}, are known *a priori*. It is also shown in this section that when applied to real-world data, where the tree-level point clouds exhibit significant variability in data quality, that this automatic optimisation routine is generally robust. However, this routine does require some tailoring to ensure consistent and accurate application across real-world data. This is achieved by empirically calibrating the approach against validation data obtained via destructive harvest. Crucially however, this empirical calibration is only required for the two parameters that are directly linked to their aforementioned theoretical counterparts. That is, these two parameters are scaled in accordance with

the input tree-level point cloud, such that this calibration does not raise concern with regards to routine transferability. Rather, this calibration stage permits application of this novel theoretical framework to real-world data.

The first stage of this automatic QSM optimisation routine exploits the possible permutations of the cover generated across the tree-level point cloud during the first step of *TreeQSM*. Figure 6.8 presented $\bar{V}$ that was derived from 20 QSMs constructed using the simulated tree-level point cloud shown in figure 6.5b for each value of d across the considered range from 0.03–0.10 m. Also presented in this figure were the 95 % confidence intervals; where it could be seen that as d trended away from d_{opt}, it was universally associated with an increase in r_{CI}.

This behaviour was explained as resulting from how consistent (or not) each of the 20 permutations of the cover were across that particular value of d. The example used was that when considering extreme scenarios, such that when the value of d is either disproportionally large or small, then each potential permutation of the cover will be highly variable. This is because the point at which the cover generation commences through the tree-level point cloud is random. As each potential permutation of the cover is significantly different from the next, then so will be the subsequent stages of segmentation and cylinder fitting. Resultantly, each estimate of tree-scale volume will by highly variable. Intrinsically then, as d trends towards d_{opt}, the variability in the potential permutations of the cover reduces. That is, as d_{opt} begins to sufficiently well partition the tree-level point cloud according to the local sections of the underlying structure, each new permutation of the cover will likely be similar.

Thus, the first stage of this automatic QSM optimisation routine is to identify the value of d that minimises variation in V; this value of d provides the initial solution of d_{opt}. This is achieved by constructing multiple QSMs across a wide set of values of d, $\{d\}$, that are guaranteed to contain d_{opt}. To ensure $\{d\}$ does indeed contain d_{opt}, the range is determined by the mean nearest neighbour distance, $\overline{d_{NN}}$, across the tree-level point cloud as:

$$\{d\}(1) = \frac{1}{2}\overline{d_{NN}} \tag{6.5}$$

$$\{d\}(N) = 2\overline{d_{NN}} \tag{6.6}$$

In the instance of the simulated tree-level point cloud illustrated in figure 6.5b, where a voxel edge length of 0.04 m was used during the downsampling, then $\{d\}(1) = 0.02\,\text{m}$ and $\{d\}(N) = 0.08\,\text{m}$. Across this range, a small step of 0.005 m would be an appropriate interval between values of d, as it was demonstrated in figure 6.8 that such a step did not result in any jumps of $\bar{V}$ lying outside the 95 % confidence intervals of the neighbouring values of d. Using this step then, $\{d\}$ for the simulated tree-level point cloud would be defined as $\{d\} = \{0.020, 0.025, 0.030, ..., 0.070, 0.075, 0.080\}$. For each value of d, 20 QSMs are constructed, in line with figure 6.7 illustrating that across 20 QSMs, the observed variation in V could be well captured. This results then, in the construction of 300 QSMs of the simulated tree-level point cloud.

For the 20 QSMs inside each parameter set, $\bar{V}$ and σ_V, are calculated to produce the corresponding sets $\{\bar{V}\}$ and $\{\sigma_V\}$. $\{\sigma_V\}$ then, indicates the dispersion of the values of V constructed by a given value of d. This could be used to identify the value of d in $\{d\}$ that yields the minimum variation in V. However, if the range across $\{\bar{V}\}$ is sufficiently large, then a comparison of $\{\sigma_V\}$ would be susceptible to bias towards smaller values of $\bar{V}$. Therefore, to mitigate this potential risk, the dimensionless coefficient of variation of the volume, CV_V, describing the variability in the spread of V, with respect to $\bar{V}$, is calculated to produce the set, $\{CV_V\}$, defined as:

$$CV_V = \frac{\sigma_V}{\bar{V}} \tag{6.7}$$

From this corresponding new set then, an initial solution of d_{opt} can be defined as:

$$d_{opt} = \{d\}(\text{argmin}\{CV_V\}) \tag{6.8}$$

That is, d_{opt} is the value of d in $\{d\}$ that minimises $\{CV_V\}$.

The second stage of this automatic QSM optimisation routine is to confirm this initial solution of d_{opt}. This is undertaken through the observation identified in figure 6.6a that when the value of d falls below d_{opt}, then it is likely that the stem will be poorly reconstructed. It could be seen in the figure that the cylinders comprising the stem appeared to be of seemingly random size and orientation. This is as would be expected as when d falls below d_{NN}, there is no longer any guarantee that there will be a sufficient number of neighbouring patches to generate any viable information determining the local

underlying surface of the tree.

This is further demonstrated in figure 6.15. In both illustrations, the section of point cloud from the simulated tree-level point cloud pictured in figure 6.5b, representing the stem from 2.0–2.7 m above the ground is shown. Overlaying this section of point cloud is the cylinder from the QSM derived from the simulated tree-level point cloud that was constructed on this particular section of point cloud. In the instance of figure 6.15a, the value of d used in the reconstruction was 0.02 m, whilst in the instance of figure 6.15b, the value of d was 0.08 m. It can be seen that a small value of d has resulted in the erroneous reconstruction of the stem, with the cylinder taking on a seemingly random radius, length and orientation. However, when a large value of d was used, the cylinder appears to appropriately resemble the underlying section of tree-level point cloud.

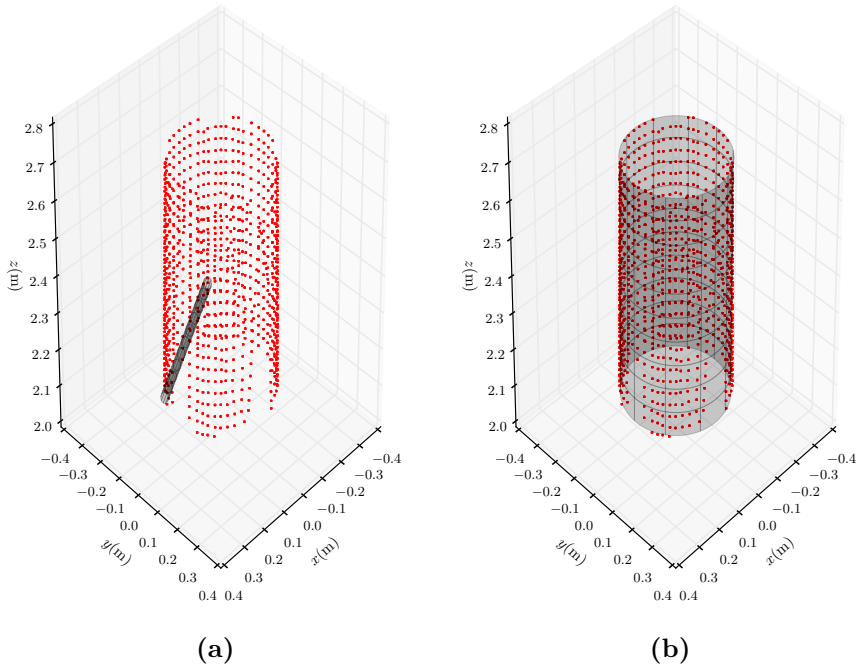


Figure 6.15: Section of point cloud taken from the simulated tree-level point cloud shown in figure 6.5b that represents the stem from 2.0–2.7 m above the ground; overlaid are the cylinders from two QSMs constructed on the tree-level point cloud that were modelled on this section of point cloud, where (a) $d = 0.02$ m and (b) $d = 0.08$ m.

From this then, the second stage of this automatic QSM optimisation routine aims to affirm the initial solution of d_{opt}, by confirming that at this value of d, the stem has been sufficiently well modelled. Per QSM, this is achieved by selecting five cylinders from along the length of the stem. The five cylinders selected are those at relative lengths of 7.5 %, 10.0 %, 12.5 % and 15.0 % along the stem. Once each of these cylinders have

been identified, their coordinates can drive a z-axis pass-through filter to extract the section of tree-level point cloud that the cylinders were supposedly modelled upon. An independent cylinder fit is then conducted on this section of point cloud via the RANSAC cylinder fitting routine described in chapter 5 section 5.2.7. The radii between the two cylinder fits, r_i and r_{QSM}, can then be compared to determine the conformity of the QSM-derived cylinder to the trunk, t_c, as the mean of the five checked cylinders, defined as:

$$t_c = \frac{1}{N}\sum_{i=1}^{N}(1 - \frac{|r_i - r_{QSM}|}{\frac{(r_i + r_{QSM})}{2}}) \tag{6.9}$$

This conformity is assessed across all QSMs inside each value of d in $\{d\}$ to create to corresponding set, $\{T_c\}$. That is, T_c is the mean value of t_c from the 20 QSMs that have been constructed per parameter set. From this corresponding new set then, the solution of d_{opt} is confirmed under the condition that:

$$\{T_c\}(d_{opt}) \approx 1 \tag{6.10}$$

That is, d_{opt} has an associated value of T_c that is roughly equal to 1 (i.e., the cylinder to trunk conformity check is indicating that the stem has been sufficiently well modelled).

To recap then, d_{opt} is determined under this automatic QSM optimisation routine by the value of d that minimises $\{CV_V\}$ and maximises $\{T_c\}$. This approach is demonstrated in figure 6.16 to correctly optimise d for the simulated tree-level point cloud shown in figure 6.5b. This simulated tree-level point cloud had a volume of $1.720\,\mathrm{m}^3$ whose $d_{opt} = 0.06\,\mathrm{m}$. Here, as aforementioned, $\{d\} = \{0.020, 0.025, 0.030, ..., 0.070, 0.075, 0.080\}$ and 20 QSMs have been constructed per value of d. In the figure, the response of $\{\bar{V}\}$, $\{CV_V\}$ and $\{T_c\}$ to $\{d\}$ is illustrated.

It can be seen from this figure, that both the minimisation of $\{CV_V\}$ and the maximisation of $\{T_c\}$ have occurred at the same value of d. At this value of d, the corresponding value $\bar{V}$ gave the best estimate of tree-scale volume closest to the actual, leading to a $0.3\,\%$ overestimate. Thus, for such simulated tree-level point clouds that were captured via near-perfect simulated sampling regimes, and exhibit a constant d_{NN} across the cloud, this theoretical routine that has been developed here produces an exact solution of d_{opt}.

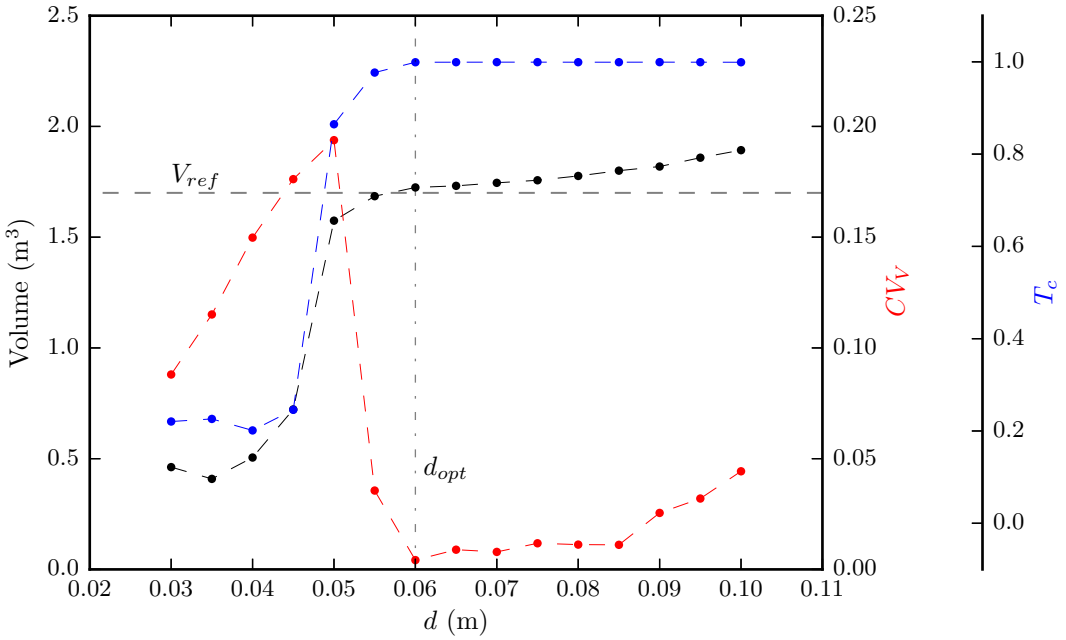


Figure 6.16: The response of $\{\bar{V}\}$, $\{CV_V\}$ and $\{T_c\}$ to $\{d\}$ for the simulated tree-level point cloud shown in figure 6.5b.

To demonstrate the appropriateness of this automatic QSM optimisation routine on real-world data, the routine has been applied to the leaf-off tree-level point cloud of AH-02 acquired from Alice Holt Forest, UK (see chapter 4 section 4.2.3). This tree-level point cloud has been selected as tree-scale AGB and average wood density has been obtained via destructive harvest, such that a reasonable estimate of tree-scale volume ($2.378\,\mathrm{m}^3$) is known *a priori* to enable pre-determination of the value of d_{opt}. Here, $\{d\} = \{0.035, 0.040, 0.045, ..., 0.0950, 0.955, 0.100\}$ with 20 QSMs constructed per value of d, such that 280 QSMs have been constructed. Figure 6.17 presents the response of $\{\bar{V}\}$, $\{CV_V\}$ and $\{T_c\}$ to $\{d\}$.

It can be seen in this figure that the response of these three sets to $\{d\}$ across real-world data does vary from the response across the simulated data. Of most prominence, it can be observed that $\{CV_V\}$ now stabilises above a certain value of d such that there is no longer any discernible minima. It can also be seen that for the value of d that produces the corresponding estimate of $\bar{V}$ closest to actual, the corresponding T_c is no longer the maxima. Both of these observations are a direct result of the introduction of variable quality data to the *TreeQSM* algorithm. That is, when this automatic QSM optimisation routine is exposed to real-world data, the value of d that minimises $\{CV_V\}$ and maximises $\{T_c\}$ no longer provides d_{opt}.

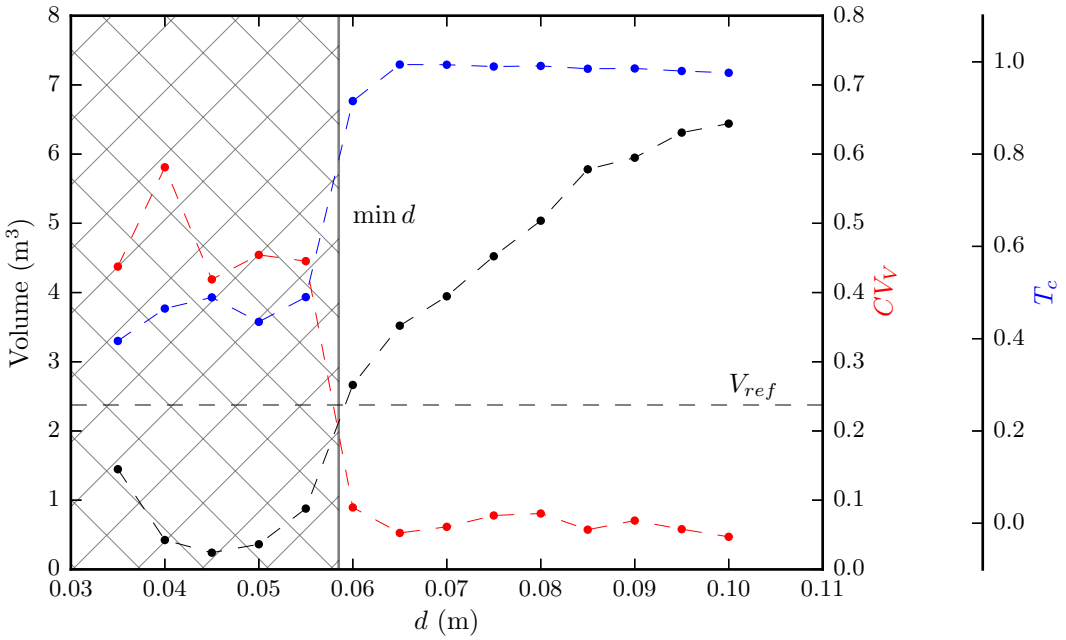


Figure 6.17: The response of $\{\bar{V}\}$, $\{CV_V\}$ and $\{T_c\}$ to $\{d\}$ for the leaf-off point cloud of AH-02 acquired from Alice Holt Forest, UK.

To compensate for this, two simplistic scalar thresholds are introduced at this stage to permit the reliable application of this automatic QSM optimisation routine to real-world data. The first threshold is the maximum permitted value of CV_V, $CV_{V\max}$, relative to the minimum observed value of CV_V, as determined by the scalar parameter S_{CV_V}, defined as:

$$CV_{V\,max} = S_{CV_V} \min\{CV_V\} \tag{6.11}$$

The second threshold is the minimum permitted value of T_c, $T_{c\min}$, relative to the maximum observed value of T_c, as determined by the scalar parameter S_{T_c}, defined as:

$$T_{c\min} = S_{T_c} \min\{T_C\} \tag{6.12}$$

That is, d_{opt} is now selected by working from the first value of d in $\{d\}$ and iterating sequentially through $\{d\}$ until encountering the first value of d conforming to both thresholds $CV_{V\max}$ and $T_{c\min}$. This has been demonstrated in figure 6.17, where $S_{CV_V} = 5$ and $S_{T_c} = 0.95$. The section of the graph that has been hatched indicates the parameter space at which these thresholds have not been met, with the remaining area of the graph

indicating the parameter space at which these thresholds have been met. As such, d_{opt} is identified as the first encountered value of d lying inside this permissible parameter space. In this instance $d_{opt} = 0.06\,\text{m}$, which coincides with the value of d that provides the best corresponding estimate of $\bar{V}$, coming in at $2.447\,\text{m}^3$, resulting in an overestimate of 5.2 %.

These two scalar values then, S_{CV_V} and S_{T_c} are empirically calibrated in section 6.3.1 using data acquired from the destructive harvest of 50 individuals post-scanning. The introduction of these two thresholds, whilst not optimal, is necessary to allow this theoretical approach for the optimisation of the QSMs to be applied to real-world data. Significantly though, these two parameters simply scale the previously theoretical maxima and minima. As such, this routine presents a significantly cleaner and more robust approach than simply calibrating d against an assortment of validation data and then universally applying that calibrated value of d to all future tree-level point clouds. That is, the automatic QSM optimisation routine presented here is based on a theoretical approach that explicitly assesses the behaviour of *TreeQSM* to each new input tree-level point cloud that it is provided.

6.2.4 Error quantification

The previous section presented a novel routine for the automatic optimisation of QSMs constructed through the *TreeQSM* algorithm. As such, the software stack combining *treeseg*, *TreeQSM* and this automatic QSM optimisation routine, realises a new capacity to routinely and near-automatically estimate tree- and plot-scale AGB from TLS data on a wide-scale for the first time. This would potentially enable TLS-derived estimates of tree- and plot-scale AGB to become viable alternatives to the allometric methods; identified in chapter 2 as in need of replacement due to their large associated uncertainties. However, for these TLS-derived estimates to become viable, it must be demonstrated that tree- and plot-scale AGB can indeed be accurately retrieved from TLS data.

This demonstration is performed in this chapter by applying the software stack to TLS data acquired from trees that have been subsequently destructively harvested to permit validation of these new methods. In all instances, post-scanning, these trees have been felled with direct measurement of tree-scale AGB undertaken through weighing. This permits an explicit quantification of the error in the TLS-derived estimates of tree-

scale AGB, $\mathrm{AGB_{TLS}}$, with the direct measurement of tree-scale AGB, AGB_{ref}. $\mathrm{AGB_{TLS}}$ is defined by the optimised QSM estimate of tree-scale volume, $\bar{V}$, and the tree-scale average wood density, ρ, as:

$$\mathrm{AGB_{TLS}} = \bar{V}\rho \tag{6.13}$$

The error in the estimate of $\mathrm{AGB_{TLS}}$ can be described, relative to AGB_{ref}, as the relative tree-scale error, ε_t, defined as:

$$\varepsilon_t = \frac{|\mathrm{AGB}_{ref} - \mathrm{AGB_{TLS}}|}{\mathrm{AGB}_{ref}} \tag{6.14}$$

Across a plot consisting of N individuals, the TLS-derived estimate of plot-scale AGB, $\mathrm{AGB}_{\mathrm{TLS},p}$, is defined through the summation of $\mathrm{AGB_{TLS}}$ across the stems, defined as:

$$\mathrm{AGB}_{\mathrm{TLS},p} = \sum_{i=1}^{N} \mathrm{AGB}_{\mathrm{TLS,i}} \tag{6.15}$$

Hence, the error in the estimate of $\mathrm{AGB}_{\mathrm{TLS},p}$ can be described, relative to the direct measurement of plot-scale AGB, $\mathrm{AGB}_{ref,p}$, as the relative plot-scale error, ε_p, defined as:

$$\varepsilon_p = \frac{|\mathrm{AGB}_{ref,p} - \mathrm{AGB}_{\mathrm{TLS},p}|}{\mathrm{AGB}_{ref,p}} \tag{6.16}$$

ε_t and ε_p then, provide explicit quantification in the error of TLS-derived estimates of tree-scale and plot-scale AGB, that can be used to demonstrate the accuracy to which tree- and plot-scale AGB can be retrieved from TLS data. Across a number of individuals, a generalisation of the tree-scale error can be generated through the mean average error of ε_t, $\bar{\varepsilon_t}$, across N individuals, defined as:

$$\bar{\varepsilon_t} = \frac{1}{N}\sum_{i=1}^{N} \frac{|\mathrm{AGB}_{ref} - \mathrm{AGB_{TLS}}|}{\mathrm{AGB}_{ref}} \tag{6.17}$$

6.3 Results

This section is broken into two distinct parts. First, section 6.3.1.1 utilises validation data acquired from 50 *Eucalyptus* spp. individuals from Rushworth State Forest, Australia,

to calibrate the developed automatic QSM optimisation routine. That is, to calibrate the scalar parameters S_{T_c} and S_{CV_V} such that, through combination of this routine with the *TreeQSM* algorithm, it is now possible to automatically retrieve tree-scale AGB from tree-level point clouds.

With this calibration complete, section 6.3.1.1 then applies these methods to the retrieval of tree-scale AGB from the same validation data acquired from Rushworth State Forest, Australia. Section 6.3.2 then applies the methods to validation data acquired from 3 *Quercus* spp. individuals from Alice Holt Forest, UK. In both of these sections, the error in the TLS-derived estimates of AGB are explicitly quantified.

6.3.1 Rushworth State Forest, Australia

This section employs TLS data acquired from 65 individuals across 2 x 0.25 ha plots (designation: RUSH06 and RUSH07) from *Eucalyptus* spp. open forest in Rushworth State Forest, Australia. These data form the basis of the proof of concept study of Calders et al. (2014), reviewed in chapter 3 section 3.2.2.2, that demonstrated, for the first time, that these new TLS-derived methods could accurately retrieve tree- and plot-scale AGB from TLS data.

The TLS data for these 65 individuals were obtained in May 2012 using data acquisition protocol identical to that used for plot KARA-001 (see chapter 4 section 4.2.1) [2]. Post-scanning, direct measurement of tree-scale AGB was obtained from the destructive harvest, and subsequent weighing of the 65 individuals [3]. Alongside direct measurement of tree-scale AGB, average wood density, ρ, was also measured by oven drying multiple disks acquired through the stems of each of the three present *Eucalyptus* species.

Chapter 5 presented the results from the application of the *treeseg* algorithm to the 3.25 ha of TLS data acquired during the work for this thesis. As the interest of chapter 7 lies with the retrieval of AGB from the 441 individual tree-level point clouds extracted from these data, here, for consistency, only the 50 individuals from these data with $D \geq 0.2$ m have been considered in this section. Additionally, in line with the requirement for the most uniformly distributed tree-level points clouds possible as input to the reconstruction process; these 50 tree-level point clouds have been downsampled via voxel grid downsampling using a voxel grid edge length of 0.04 m. The tree-level point clouds of these 50 individuals from the plots RUSH06 and RUSH07 are presented in figures 6.18a and 6.18b respectively.

[2]The TLS data were not collected by the author. These data are owned by Wageningen University, 6708 PB Wageningen, Netherlands (Contact: Dr. Kim Calders, kim.calders@ucl.ac.uk). Permission was granted to use these data in this thesis.

[3]The destructive harvest data were not collected by the author. These data are owned by The University of Melbourne, Department of Forest and Ecosystem Science, Water Street, Creswick, Victoria 3363, Australia (Contact: Dr. Chris Weston, weston@unimelb.edu.au). Permission was granted to use these data in this thesis.

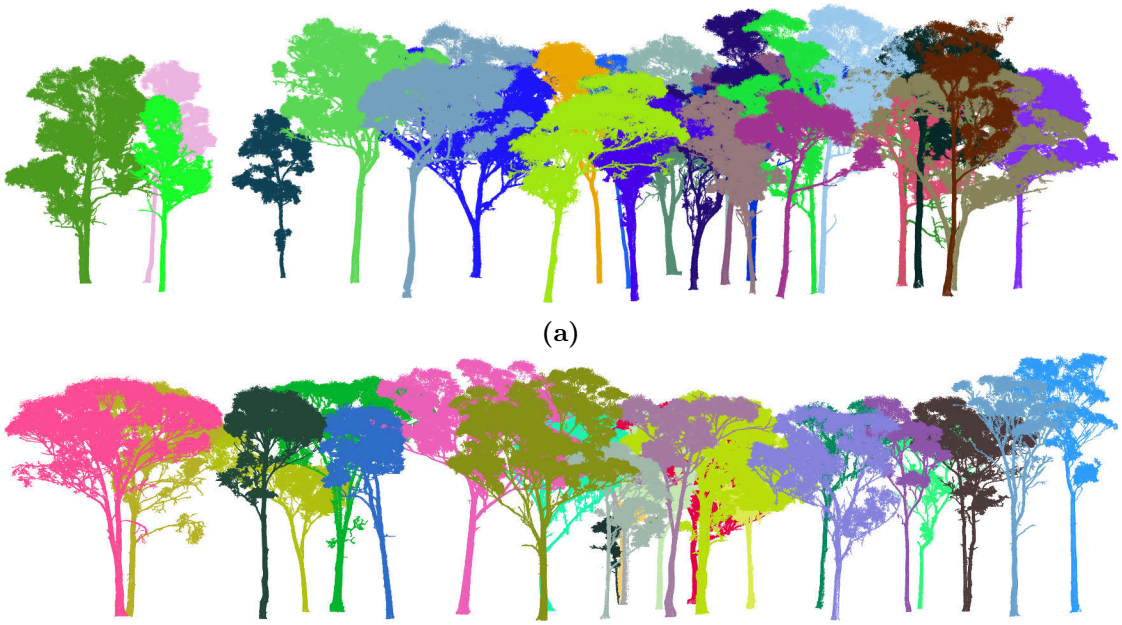


Figure 6.18: The 50 tree-level point clouds forming the validation dataset considered here from Rushworth State Forest, Australia, across the plots: (a) RUSH06 and (b) RUSH07; the individuals shown here have $D \geq 0.2\,\text{m}$ and the tree-level point clouds have been downsampled via voxel grid downsampling using a voxel edge length of 0.04 m.

6.3.1.1 Calibration of the automatic QSM optimisation routine

The scalar parameters, S_{T_c} and S_{CV_V}, permitting application of the theoretical automatic QSM optimisation routine to real-word data, are calibrated here against the validation data of the 50 trees from RUSH06 and RUSH07. To recap, S_{T_c} is the parameter defining the maximum permissible non-conformity of the radius of the cylinders inside each QSM representing the trunk, to the corresponding independently derived radius of the trunk; scaled with respect to the specific tree-level point cloud in question. S_{CV_V} is the parameter defining the maximum permissible coefficient of variation of the volume, across each of the QSMs in each considered parameter set, also scaled with respect to the specific tree-level point cloud in question.

As the accurate reconstruction of the stem is paramount to the accurate retrieval of tree-scale volume, S_{T_c} is calibrated first. This calibration is undertaken through application of the automatic QSM optimisation routine to the 50 tree-level point clouds, using a range of values of S_{T_c} from 0.700–1.00 , with a step of 0.025 For each tree-level point cloud, the set, $\{d\}$, comprising the considered values of the input parameter for optimisation, d, was defined as $\{d\} = \{0.040, 0.045, 0.050, ..., 0.140, 0.145, 0.150\}$, whilst

N was held constant at 3. Per tree-level point, 20 QSMs were constructed for each considered value of d, resulting in the construction of a total of 460 QSMs. As the same voxel grid edge length was used in the sampling as with the tree-level point cloud shown in figure 6.11b, the height-resolved mean nearest neighbour distance distribution was near-identical to that shown in figure 6.12. As such, in line with this height-resolved mean nearest neighbour distance across the tree-level point clouds, all cylinders in the QSMs with radius less than 0.025 m were pruned. As explained in the methods section, this discarding of cylinders is undertaken because cylinders of a radius below this threshold have been inferred from ambiguous point arrangements.

For each considered value of S_{T_c}, $\mathrm{AGB}_{\mathrm{TLS}}$, ε_t, $\bar{\varepsilon_t}$ and $\mathrm{AGB}_{\mathrm{TLS},p}$ were calculated. The response of $\mathrm{AGB}_{\mathrm{TLS},p}$ to the considered range of values of S_{T_c} is illustrated in figure 6.19. In this figure, the intervals placed around $\mathrm{AGB}_{\mathrm{TLS},p}$ are $\bar{\varepsilon_t}$.

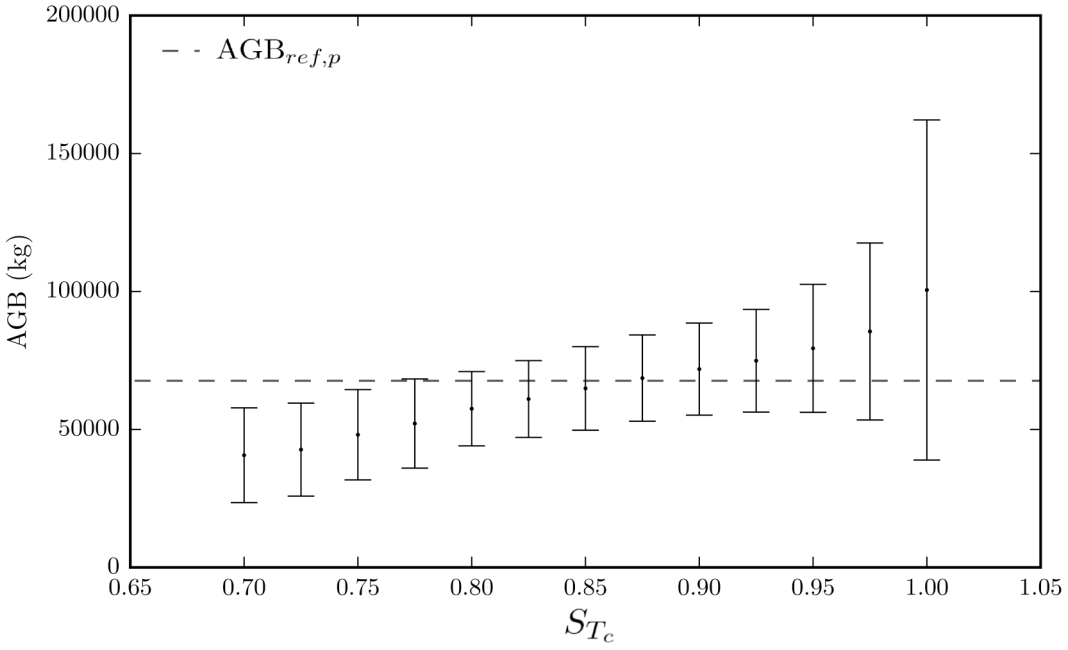


Figure 6.19: The response of $\mathrm{AGB}_{\mathrm{TLS},p}$ to the considered values of the scalar parameter, S_{T_c}, defining the maximum permissible non-conformity of the radius of the cylinder inside each QSM representing the trunk, to the corresponding independently derived radius of the trunk; the intervals shown about $\mathrm{AGB}_{\mathrm{TLS},p}$ are $\bar{\varepsilon_t}$.

It can be seen that across the range of values of S_{T_c}, $\mathrm{AGB}_{\mathrm{TLS},p}$ ranges from 40 655–100 535 kg, whilst $\bar{\varepsilon_t}$ ranges from 22.8–61.3 %. The value of S_{T_c} that minimises ε_p is 0.875 , whilst the value that minimises $\bar{\varepsilon_t}$ is 0.900 As the interest here lies with producing the most accurate estimates of tree-scale AGB, the calibrated value of S_{T_c} is defined as 0.900 That is, although $S_{T_c} = 0.875$ produced the best plot-scale estimate of AGB, $S_{T_c} = 0.900$

best reduced the mean observed error across the 50 individuals.

With S_{T_c} calibrated, the second required step is to calibrate the parameter S_{CV_V}, whilst holding $S_{T_c} = 0.900$. In a similar fashion to the calibration of S_{T_c}, the automatic QSM optimisation routine is applied to the 50 tree-level point clouds using a range of values of S_{CV_V} from 1–6 , with a step of 0.025 The response of $\mathrm{AGB}_{\mathrm{TLS},p}$ to the considered range of values of S_{CV_V}, can be seen in figure 6.20. In this figure, the intervals placed around $\mathrm{AGB}_{\mathrm{TLS},p}$ are $\bar{\varepsilon_t}$.

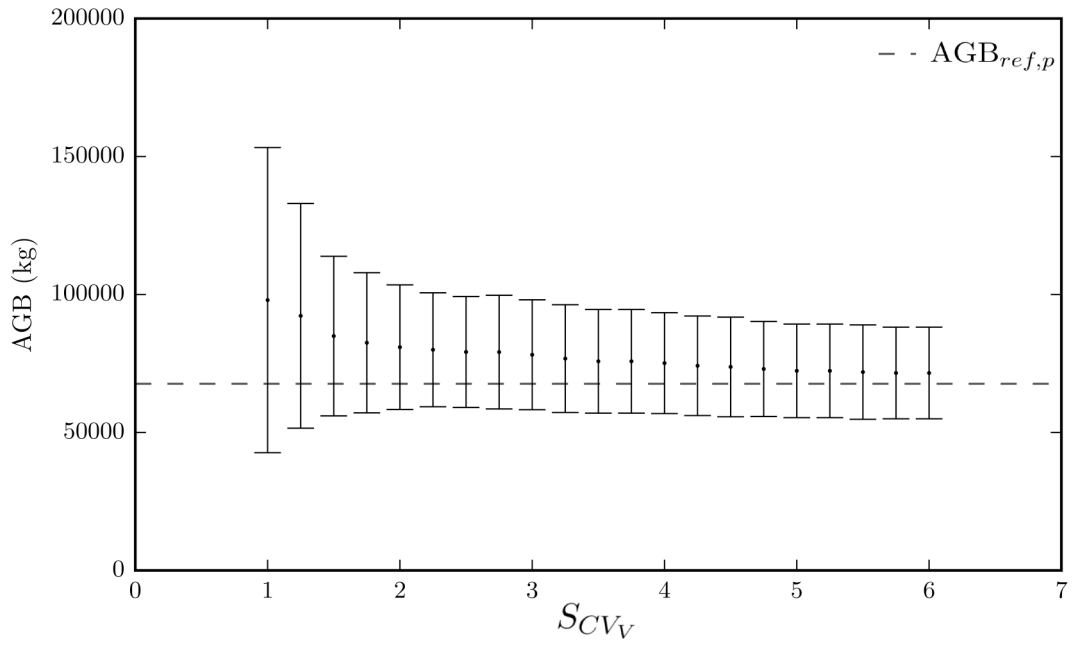


Figure 6.20: The response of $\mathrm{AGB}_{\mathrm{TLS},p}$ to the considered values of the scalar parameter, S_{CV_V}, defining the maximum permissible coefficient of variation of the volume across the QSMs in each considered parameter set, the intervals shown about $\mathrm{AGB}_{\mathrm{TLS},p}$ are $\bar{\varepsilon_t}$.

It can be seen that across the range of considered values of S_{CV_V}, $\mathrm{AGB}_{\mathrm{TLS},p}$ ranges from 71 547–97 952 kg, whilst $\bar{\varepsilon_t}$ ranges from 23.2–56.6 %. As S_{T_c} has been held constant at 0.9 , above a value of $S_{CV_V} = 5$, the value of both ε_p and $\bar{\varepsilon_t}$ reach a minimum. Thus, the calibrated value of S_{T_c} is defined as 5.

Figure 6.21 presents the optimised QSMs for each of the considered 50 tree-level point clouds that have been automatically reconstructed using the routine where $S_{T_c} = 0.9$ and $S_{CV_V} = 5$. Per tree, the QSM shown is the one in the set of 20 across the optimised value of d, whose volume lies closest to the mean volume.

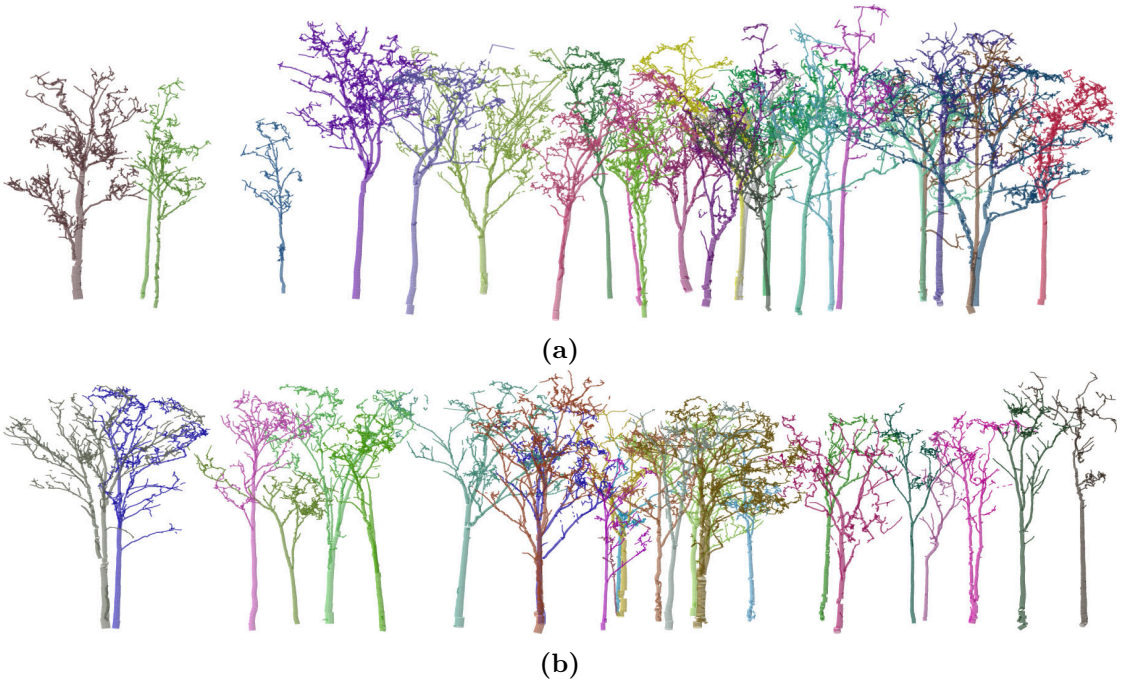


Figure 6.21: The 50 optimised QSMs constructed from the tree-level point clouds in figure 6.18 from (a) RUSH06 and (b) RUSH07; in each QSM, the cylinders below a radius of 0.025 m have been pruned in accordance with the height-resolved mean nearest neighbour distance across the tree-level point cloud.

6.3.1.2 Error quantification

With the calibration of the automatic optimisation routine complete, it is now possible to routinely, and entirely automatically, retrieve tree-scale AGB from tree-level points clouds. This section explicitly quantifies the errors in doing so across the 50 tree-level point clouds from RUSH06 and RUSH07.

For each tree-level point cloud, the automatic QSM optimisation routine considers the set $\{d\} = \{0.040, 0.045, 0.050, ..., 0.140, 0.145, 0.150\}$, whilst holding N constant at 3. As dictated by the previous section, the scalar parameters are defined as $S_{CV_V} = 5$ and $S_{T_c} = 0.95$. 20 QSMs were constructed for each considered value of d, resulting in the construction of a total of 460 QSMs. In line with the height-resolved mean nearest neighbour distance across the tree-level point clouds, all cylinders in the QSMs with radius less than 0.025 m were pruned.

Figure 6.22 presents the comparison between the direct measurement of tree-scale AGB and the automatically derived TLS estimates of tree-scale AGB across the 50 trees considered from plots RUSH06 and RUSH07. Accompanying the TLS-derived estimate of tree scale AGB are the 95 % confidence intervals, derived from the 20 QSMs cor-

responding to the optimised value of d, in accordance with equation 6.3. The figure demonstrates, that at the tree-scale, there is a general trend for the TLS-derived estimates of tree-scale AGB to be an overestimate. Across the 50 trees, $\mathrm{AGB}_{\mathrm{ref},p} = 67\,632\,\mathrm{kg}$, whilst $\mathrm{AGB}_{\mathrm{TLS},p} = 72\,997\,\mathrm{kg}$. From these results then, $\varepsilon_p = 7.9\,\%$ and $\bar{\varepsilon_t} = 23.3\,\%$ That is, at the tree-scale, the mean observed error was 23.3 %, which at the plot-scale, reduced down to a 7.9 % overestimate. Finally, it can be seen that apart from 6 outliers, including three of the smallest trees, with AGB $< 500\,\mathrm{kg}$, the remaining 44 confidence intervals, constructed about the variation in the tree-scale volume of the 20 optimised QSMs, contain the value of tree-scale AGB from the direct measurement. That is, these confidence intervals appear to reasonably capture the uncertainty in the estimate of tree-scale AGB, U_t, across these data.

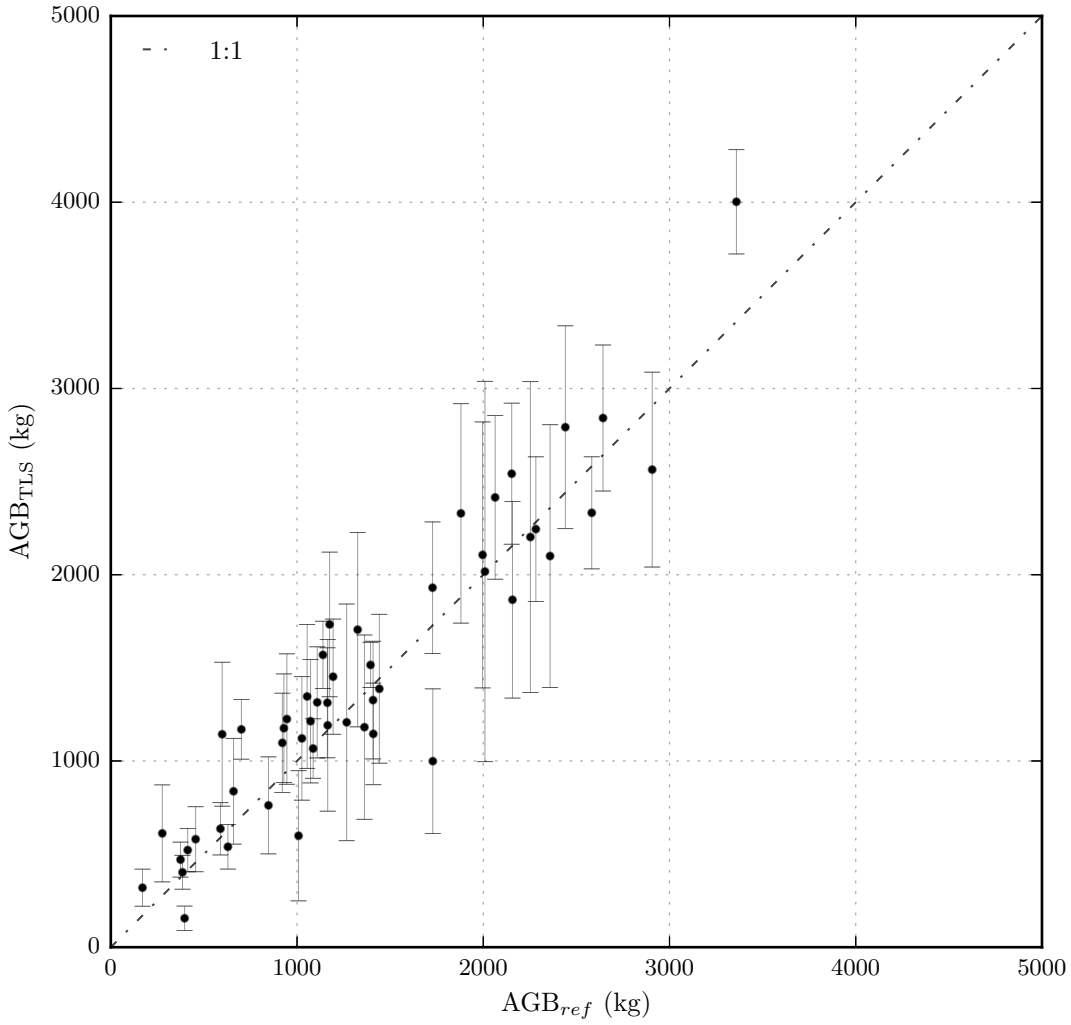


Figure 6.22: Comparison between the direct measurement and TLS-derived estimates of tree-scale AGB across the plots RUSH-06 and RUSH-07.

These TLS-derived estimates of tree- and plot-scale AGB are then compared with their allometric counterparts. This has been undertaken by applying the NLS pan-tropical allometric model developed in chapter 2 and applying it to these 50 individuals. Despite this plot not being tropical forest, the application of this model is suitable here as the allometric dataset of Chave et al. (2014) contains 46 harvested *Eucalyptus* spp. individuals from open forest in Australia. To recount, this allometric model employs a single amalgamated allometric regressor, that takes the form:

$$\text{AGB} = 0.179(D^2 H \rho)^{1.102} + 80.235 \tag{6.18}$$

For each considered tree, the regressor D is obtained from the tree-level point cloud via the RANSAC cylinder fit method presented in chapter 5 section 5.3.8. The regressor H is obtained from the tree-level point cloud via the Euclidean distance between the minimum and maximum z-coordinates through the cloud.

Figure 6.23 presents the comparison between the direct measurement of tree-scale AGB and the allometric estimates of tree-scale AGB across the 50 trees considered from plots RUSH06 and RUSH07. Accompanying the allometric-derived estimate of tree-scale AGB is the tree-level allometric uncertainty. This uncertainty is defined through the methods that were presented in chapter 2. That is, the two additive contributors to U_t, the tree-scale allometric model uncertainty, and the tree-scale measurement uncertainty, are derived from non-linear quantile regression and Monte Carlo simulation respectively. Across the 50 trees, $\text{AGB}_{\text{ref},p} = 67\,632\,\text{kg}$, whilst $\text{AGB}_{\text{Allom},p} = 53\,870\,\text{kg}$. From these results then, $\varepsilon_p = 20.3\,\%$ and $\varepsilon_t = 20.5\,\%$. At the tree-scale then, the mean observed error was 20.4 %, which at the plot-scale, marginally reduced down to a 20.3 % underestimate.

Next, figure 6.25 compares the TLS- and allometric-derived estimates of tree-scale AGB against each other. At the plot-scale, with $\text{AGB}_{\text{TLS},p} = 72\,997\,\text{kg}$ and $\text{AGB}_{\text{Allom},p} = 53\,870\,\text{kg}$, this represented a very significant 30.2 % percentage difference between the estimates provided by the two methods.

Finally, figure 6.25 compares the distribution of the absolute errors across both the TLS- and allometric-derived estimates of tree-scale AGB. It can be seen that the absolute errors in the TLS-derived estimates of tree-scale AGB appear homoscedastic. Whereas

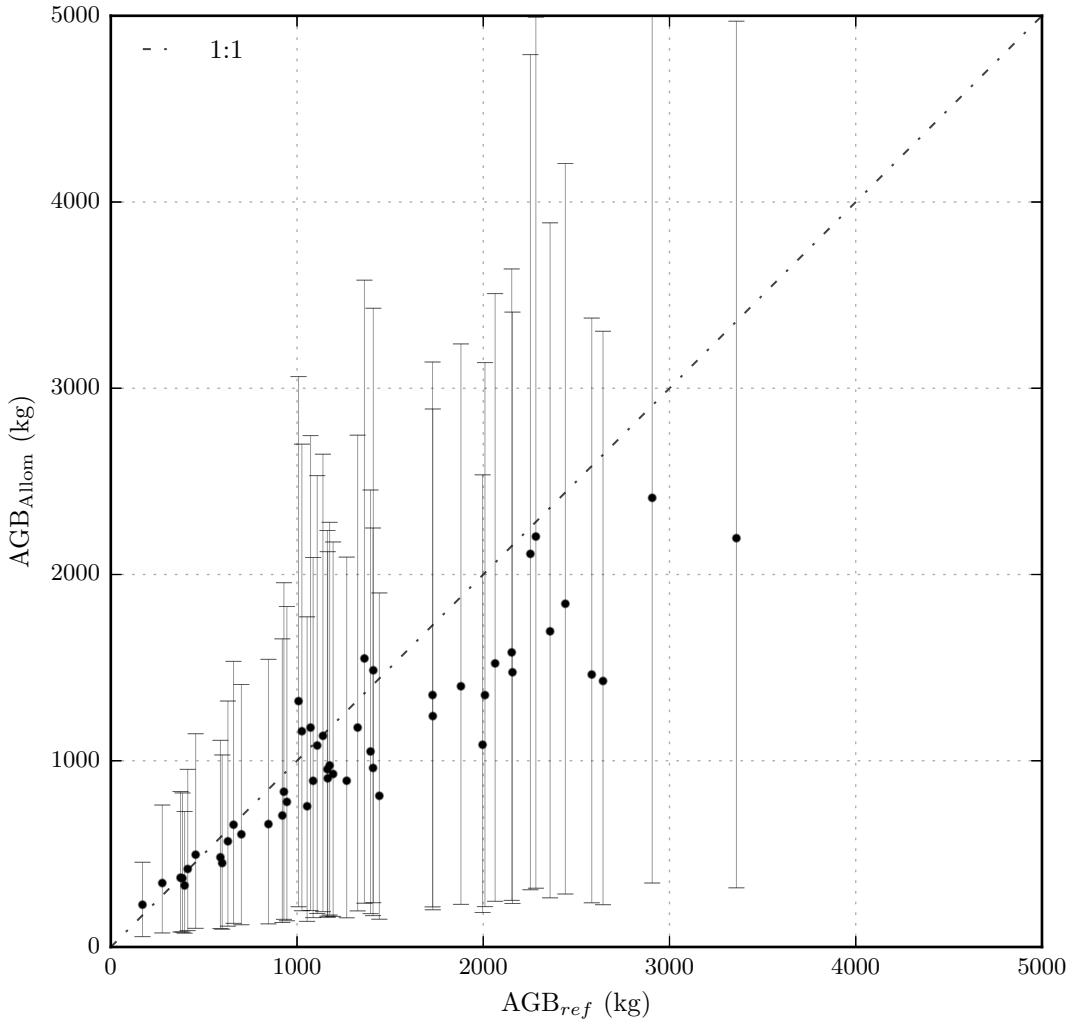


Figure 6.23: Comparison between the direct measurement and allometric-derived estimates of tree-scale AGB across the plots RUSH-06 and RUSH-07.

the absolute errors in the allometric estimates of tree-scale AGB appear correlated with the absolute value of AGB. That is, across the AGB range of 0–1000 kg, the absolute errors in the allometric estimates remain below those of the TLS-derived estimates. However, above this range, the absolute errors in the allometric estimates appear to grow exponentially.

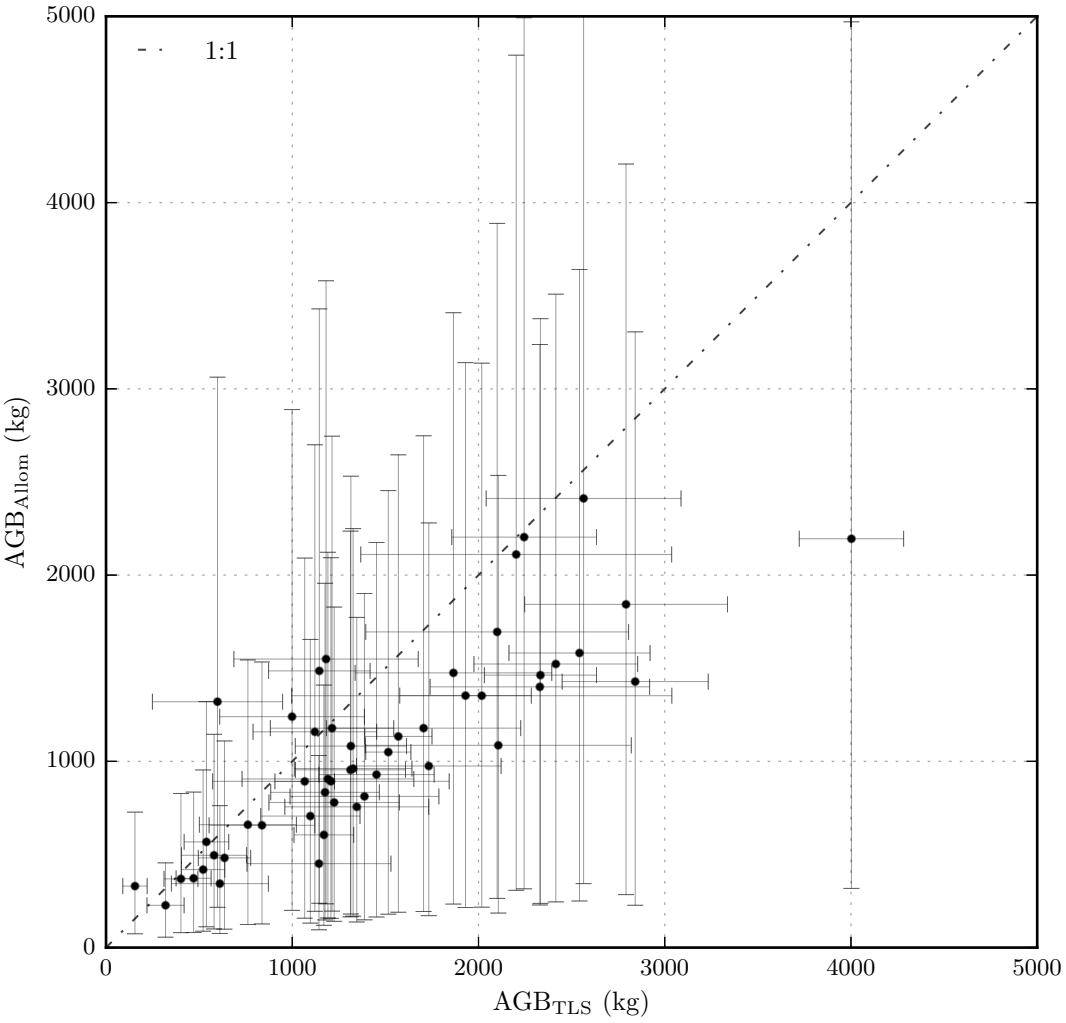


Figure 6.24: Comparison between the TLS- and allometric-derived estimates of tree-scale AGB across the plots RUSH-06 and RUSH-07.

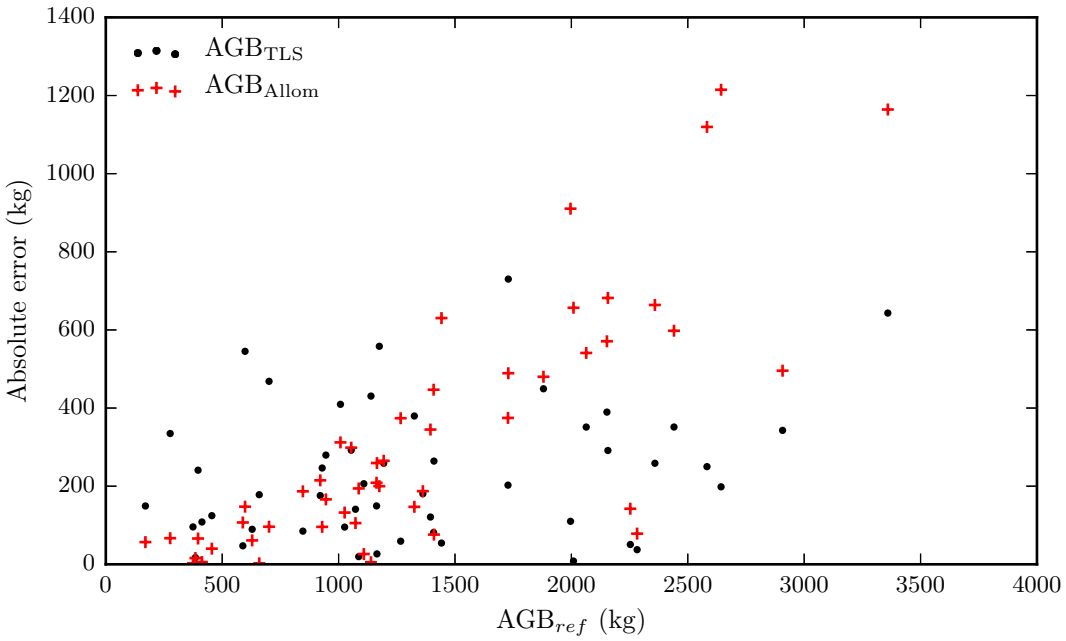


Figure 6.25: Absolute error of the TLS- and allometric-derived estimates of tree-scale AGB with the direct measurement of tree-scale AGB.

6.3.2 Alice Holt Forest, UK (AH)

This section employs the TLS data acquired from 3 *Quercus* spp. individuals from *Quercus* spp. woodland in Alice Holt Forest, UK, as described in chapter 4 section 4.2.3. The TLS data obtained across these 3 individuals were collected in April 2014 and July 2014 to sample these trees in both leaf-off and leaf-on conditions. Post-scanning, direct measurement of tree-scale AGB was obtained from the destructive harvest, and subsequent weighing of the 3 individuals [4]. Alongside direct measurement of tree-scale AGB, average wood density, ρ, was also measured by oven drying multiple disks acquired through the stems of each of the three individuals.

The data have been used in this section to serve three purposes. The first purpose is to independently validate the calibration of the automatic QSM method (i.e., through application of the routine on validation data that it was not calibrated upon). The second purpose is to explicitly quantify the error in these automatically derived estimates of tree-scale AGB. Finally, these TLS data have a unique capacity to explicitly quantify one of the underlying sources of the error in TLS-derived estimates of AGB, namely the impact of leaf returns in the tree-level point cloud, on the goodness of the reconstruction of woody structure.

For consistency, each of the tree-level point clouds have been downsampled via voxel grid downsampling using a voxel grid edge length of 0.04 m. For each tree-level point cloud, the automatic QSM optimisation routine considers the set $\{d\} = \{0.040, 0.045, 0.050, ..., 0.140, 0.145, 0.150\}$, whilst holding N constant at 3. As dictated by section 6.3.1.1, the scalar parameters are defined as $S_{CV_V} = 5$ and $S_{T_c} = 0.95$. 20 QSMs were constructed for each considered value of d, resulting in the construction of a total of 460 QSMs. In line with the height-resolved mean nearest neighbour distance across the tree-level point clouds, all cylinders in the QSMs with radius less than 0.025 m were pruned.

Figure 6.26 presents the optimised QSMs for each of the 3 trees overlaid on the tree-level point clouds from which they were constructed. The QSM shown is the one in the set of 20 across the optimised value of d, whose volume lies closest to the mean volume.

[4] The destructive harvest data were not collected by the author. These data are owned by Forest Research, Alice Holt Lodge, Farnham, Surrey, GU10 4LH, UK (Contact: Dr. Eric Casella, eric.casella@forestry.gsi.gov.uk). Permission was granted to use these data in this thesis.

The top row of images presents the leaf-off point clouds and QSMs, whilst the bottom row of images presents the leaf-on point clouds and QSMs.

Tables 6.1 and 6.2 present AGB_{ref}, $\mathrm{AGB}_{\mathrm{TLS}}$ and ε_t across each of trees in leaf-off and leaf-on conditions respectively. It can be seen that in leaf-off conditions, across the 3 trees, the error in the TLS-derived AGB leads to a 3.7 % overestimate. In leaf-on conditions, across the 3 trees, this error in the TLS-derived AGB rises to a 26.2 % overestimate. That is, across these considered data, the interference of leaf returns on the retrieval of tree-scale AGB has added 520 kg to the total TLS-derived estimate.

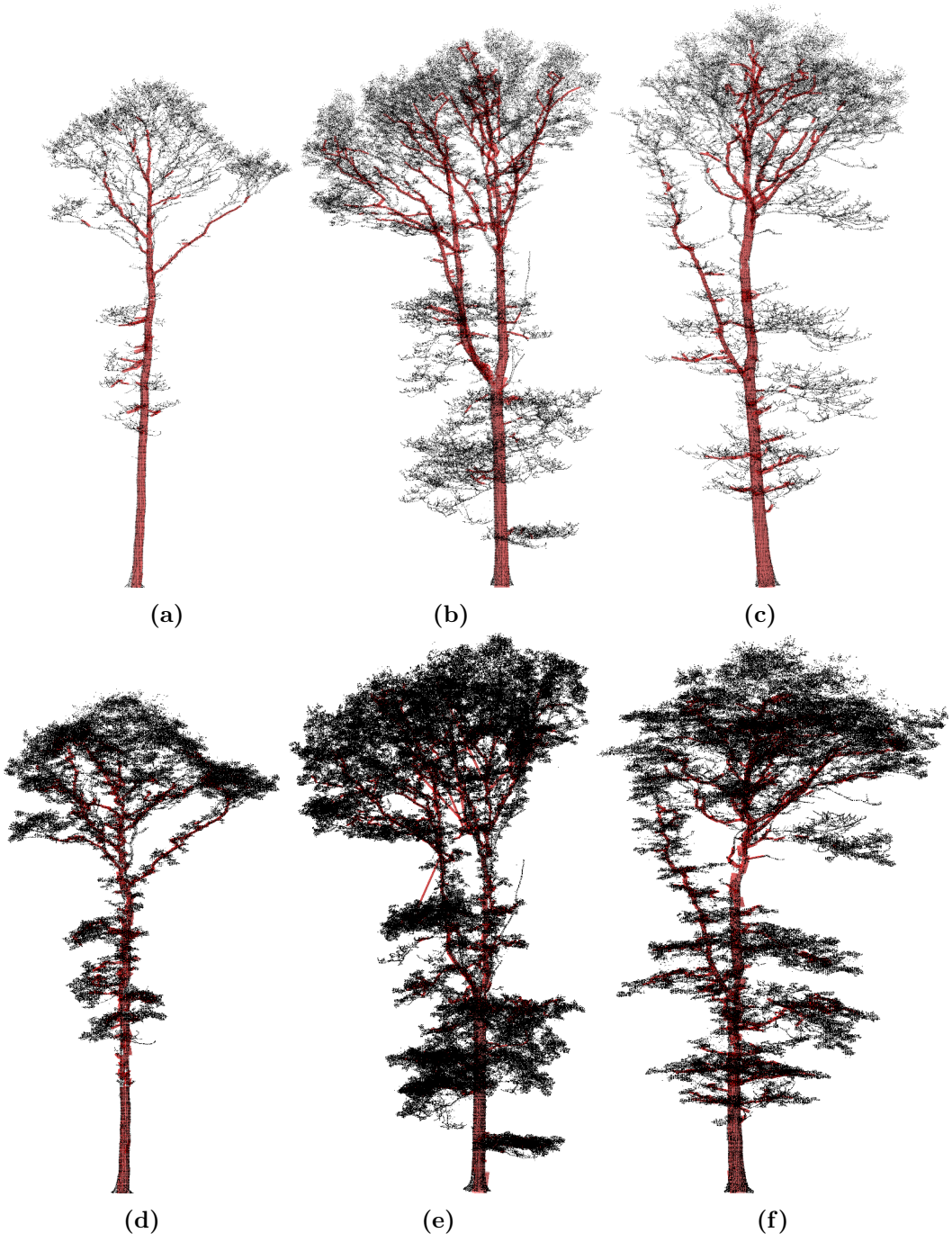


Figure 6.26: The QSMs (red) underlying tree-level point clouds (black) from which they were constructed from for the three 3 *Quercus* spp. individuals from Alice Holt Forest, UK; the upper row presents the QSMs generated from the leaf-off tree-level point clouds and the lower row presents the QSMs generated from the leaf-on tree-level point clouds.

Tree-code	AGB_{ref} (kg)	AGB_{TLS} leaf-off (kg)	ε_t (%)
AH-01	434	435	0.2
AH-02	1242	1309	5.2
AH-03	638	656	2.8
Total	2314	2400	3.7

Table 6.1: Comparison between the direct measurement and TLS-derived estimate of tree-scale AGB across the three *Quercus* spp. individuals in leaf-off conditions.

Tree-code	AGB_{ref} (kg)	AGB_{TLS} leaf-on (kg)	ε_t (%)
AH-01	434	588	26.0
AH-02	1242	1426	12.9
AH-03	638	906	42
Total	2314	2920	26.2

Table 6.2: Comparison between the direct measurement and TLS-derived estimate of tree-scale AGB across the three *Quercus* spp. individuals in leaf-on conditions.

6.4 Discussion

This chapter commenced by introducing the Raumonen et al. (2013) shape-fitting algorithm that enables the retrieval of tree-scale volume from tree-level point clouds via the reconstruction of woody structure. It was demonstrated that the accuracy of the retrieved estimate of tree-scale volume from this algorithm is sensitive to the choice of the user-defined input parameters. Currently, end-users of this algorithm select the input parameters through visual inspection, a process which is both liable to bias, and impractical on a wide-scale. As such, the remaining knowledge gap in the automatic estimation of tree- and plot-scale AGB from TLS data on a wide-scale is the automatic selection of the optimum values of these user-defined input parameters.

This chapter has presented a novel automatic QSM optimisation routine, whose approach is theoretically derived from the underlying methods of the Raumonen et al. (2013) algorithm. This approach exploits the variable behaviour of the algorithm to the ingested point cloud, to assess which input parameters simultaneously best minimise the variance observed in the tree-scale estimate of volume, and ensure a proper reconstruction of the stem. This approach then, tailors the input parameters to each specific tree-level point cloud. It was shown that when this routine was applied to perfectly sampled data, obtained via simulation, the input parameters corresponded directly with theoretical maxima and minima.

However, it was also demonstrated that when this routine was applied to real-world

data, these theoretical values are susceptible to noise. This is due either to an incomplete sampling, or a non-uniform distribution of points through the tree-level point cloud. As a result of this, two parameters were introduced into this routine to permit application of this theoretical approach to real-world data. These parameters are directly linked to their corresponding theoretical counterparts, but permit a relaxation to account for variable data quality, as scaled by the composition of the tree-level point cloud itself. That is, this routine permits the optimisation of QSMs from tree-level point clouds of variable quality, making the routine broadly relevant to TLS data acquired across differing instruments and forest types. This is in contrast to direct calibration of the user-defined input parameters though destructive harvest (i.e., forcing the QSM to give the correct value of tree-scale volume). The applicability of such methods to other TLS datasets, or, even to tree-level point clouds from the same plots is highly questionable.

Combination of the Raumonen et al. (2013) shape-fitting algorithm with this new automatic QSM optimisation routine, represents the bridging of the last remaining knowledge gap that is prohibiting the capacity to automatically retrieve tree- and plot-scale AGB from TLS data. The accuracy of this new capacity to automatically estimate tree- and plot-scale AGB from tree-level point clouds has been assessed in this chapter on destructive harvest data acquired from 53 individuals. Across these 53 individuals, post-scanning, direct measurement of AGB has been obtained via weighing. This has permitted the explicit quantification of error in these new TLS-derived methods.

Across the validation data from Rushworth State Forest, Australia, it was demonstrated, that the error in tree- and plot-scale AGB was 23.3 % and 7.9 % respectively. In almost all cases, the TLS-derived estimates resulted in an overestimate of AGB. Most notably, the reduction in the error associated with the transition from the tree to plot-scale, clearly demonstrates that the TLS-derived estimate of tree-scale AGB are broadly normally distributed. That is, the expectation is that as the number of TLS samples (i.e., number of trees in a plot) tends towards to infinity, then the error tends to zero. This is clearly demonstrated in figure 6.25, where it can be seen that the absolute errors in the TLS-derived estimates of tree-scale AGB were homoscedastic.

This is in significant contrast to the allometric-derived estimates of tree- and plot-scale AGB across the same data, where the error in tree- and plot-scale AGB was 20.4 % and20.3 % respectively. In almost all cases, the allometric estimates resulted in an un-

derestimate of AGB. Of most importance, the absence in the reduction of error when transitioning from the tree to the plot-scale is indicative of bias. This is also reflected in figure 6.25, where it can be seen that the absolute error in the allometric estimates of tree-scale AGB were heteroscedastic (i.e., correlated to AGB).

In summary, across these data from Rushworth State Forest, Australia, the direct measurement of AGB was 67 632 kg, the TLS-derived estimate of AGB was 72 997 kg and the allometric-derived estimate of AGB was 53 870 kg. This represents a percentage difference between the direct measurement and the TLS- and allometric-derived estimates of 7.9 % and 20.3 % respectively. The percentage difference between the two estimates was 30.2 %. It can be concluded then, that across the considered data, the TLS estimation of plot-scale AGB is far superior to the allometric estimate. This is clear demonstration that TLS-derived methods are indeed a viable alternative to allometrics.

The validation data from Alice Holt Forest, UK also provided validation of these new TLS-derived methods. These data were particularly valuable as they offered an independent assessment of the automatic QSM optimisation routine. That is, it was demonstrated that the two parameters introduced into the routine to permit application to real-world data are indeed readily transferable across forest types. Across these data, it was shown that the error in TLS-derived tree-scale AGB was 3.7 % and 26.2 % in leaf-off and leaf-on conditions respectively. This provides a unique insight into one of the many contributing sources of these errors, namely the interference of leaf returns on the reconstruction of woody structure. The sources of these errors, captured implicitly by the validation data considered in this chapter, are considered further in chapter 7.

6.5 Conclusions

Chapter 3 concluded by stating the knowledge gaps that currently prohibit TLS from becoming a viable, more accurate alternative to allometric methods in the estimation of tree- and plot- scale AGB. Chapter 5 presented *treeseg*, a novel algorithm developed during the work for this thesis to near-automatically extract tree-level points clouds from plot-level point clouds, so as to bridge the first of these knowledge gaps.

This chapter has presented a novel algorithm to automatically optimise the shape-fitting algorithm of Raumonen et al. (2013), to permit the automatic retrieval of tree-scale AGB from tree-level point clouds, bridging the second of the knowledge gaps. That

is, the software stack comprising the *treeseg* algorithm, the Raumonen et al. (2013) shape-fitting algorithm and this QSM optimisation routine, represent a new capacity to near-automatically estimate tree- and plot-scale AGB from TLS data.

The final knowledge gap is to demonstrate that this new capacity is indeed capable of accurately estimating tree- and plot-scale AGB, beyond the accuracies of that permitted by allometry, such that TLS can be considered a viable alternative to the allometrics. This demonstration has been performed in this chapter through application of the software stack to TLS data acquired from 53 individuals, that have subsequently been destructively harvested to obtain a direct measurement of tree-scale AGB. This direct measurement allows the explicit quantification of the error in these new TLS-derived estimates of tree-scale AGB.

It has been shown, that across these data, the errors in TLS-derived estimates of tree- and plot-scale AGB were 23.3 % and 7.9 % respectively. In comparison, it was shown that across these same data, the errors in the allometric estimates of tree- and plot-scale AGB were 20.4 % and 20.3 % respectively. It was further demonstrated that the errors in the TLS-derived estimates could be approximated as being normally distributed, explaining the observed reduction in error when transitioning from the tree to the plot-scale. In contrast, it was shown that the errors in the allometric-derived estimates were heteroscedastic, indicative of bias in the estimates. As a result, across these data, it has been demonstrated that the TLS-derived estimates of plot-scale AGB are far superior to their allometric counterparts, such that it can be concluded that these new methods are indeed viable alternatives to allometric methods.

Chapter 7 will apply these new methods to TLS data acquired during the work for this thesis from 3.25 ha of predominately tropical forest. From these data, TLS-derived estimates of tree-scale AGB for 441 individuals, and plot-scale AGB for four plots will be retrieved. These estimates will be compared with their allometric counterparts and a new pan-tropical allometric model will be presented.

7 Estimates of above-ground biomass from terrestrial laser scanning

7.1 Introduction

Chapter 2 demonstrated through a novel experimental study that the relative uncertainties in allometric predictions of pan-tropical tree- and plot-scale above-ground biomass (AGB) exceeded 80 % and 25 % respectively. As a consequence of these large uncertainties, it was shown that pan-tropical allometric-derived estimates of AGB for large swathes of tropical forest will be statistically indistinguishable from one another. As a result, it was concluded that there is a requirement for the introduction of alternative methods.

Chapter 3 reviewed the alternative method considered in this thesis, terrestrial laser scanning (TLS). It was shown in the review, that TLS data acquired from commercially available instruments, combined with new modelling techniques fitting geometric primitives to tree-level point clouds, are capable of faithfully reconstructing woody tree structure (Raumonen et al. 2013; Hackenberg et al. 2015). These quantitative structural models (QSMs) provide estimates of tree-scale volume, that married with estimates of average wood density, permit the retrieval of tree-scale AGB from TLS data. A review of the two proof-of-concept studies that applied these TLS-derived methods to simulated and real-world validation datasets, demonstrated that tree-scale volume and AGB could be accurately derived from these new methods (Disney et al. 2012; Calders et al. 2015). These new methods then, free from the assumptions that burden their allometric counterparts with such large uncertainties, present a potential paradigm shift in the way that forests are measured.

Chapter 3 concluded by identifying the knowledge gaps that restrict these new methods from being applied on a wide-scale (here meaning across tens of 1 ha forest plots), currently preventing them from becoming a viable alternative to allometrics. The first knowledge gap, the automatic extraction of tree-level point clouds from plot-level point clouds, was bridged during the work for this thesis by the development of the *treeseg* algorithm described in chapter 5. The second knowledge gap, the automatic optimisa-

tion of the QSMs, was addressed during the work for this thesis in chapter 6 by the development of a routine to automatically optimise the QSMs constructed by the shape-fitting algorithm of Raumonen et al. (2013). The final knowledge gap, the quantification of uncertainties attributed to these new TLS-derived estimates of tree- and plot-scale AGB, was assessed in chapter 6 through application of the methods to validation data acquired via destructive harvest. It was shown that across the 53 considered individuals, the relative error in tree- and plot-scale estimates of AGB was 23.3 % and 7.9 % respectively. These errors then, are significantly lower than their corresponding uncertainties encountered in the allometric estimates of tree- and plot-scale AGB.

That is, the work presented in this thesis has bridged the three knowledge gaps identified in chapter 3, such that, for the first time, there is a capacity to routinely and near-automatically retrieve estimates of tree- and plot-scale AGB from TLS data. This capacity is derived from the software stack comprising the *treeseg* algorithm, the shape fitting algorithm of Raumonen et al. (2013) and the automatic QSM optimisation routine.

This chapter demonstrates this capacity by applying the software stack to the TLS data described in chapter 4, acquired during the work for this thesis from 3.25 ha of predominantly tropical forest. From these data, tree- and plot-scale estimates of AGB are retrieved for 441 individuals, and directly compared with their corresponding allometric estimates. Finally, this chapter concludes by presenting a new pan-tropical allometric equation that has been constructed by augmenting the Chave et al. (2014) allometric dataset with these new TLS data.

7.2 Methods

Section 7.2.1 briefly recounts the methods developed in chapters 4, 5 and 6 that permit the retrieval of tree- and plot-scale AGB from TLS data. Section 7.2.2 then briefly recounts the methods developed in chapter 2 for the allometric prediction of tree- and plot-scale AGB, and associated uncertainty.

7.2.1 TLS-derived estimates of tree- and plot-scale AGB

As reviewed in chapter 4 section 4.2, TLS data were acquired from 3 x 1 ha censused plots in moist, Terra Firme, lowland, mixed species, old-growth tropical forest in Lopé National Park, Gabon (LPG-01), Caixuanã National Forest, Brazil (CAX-A) and Nouragues Na-

ture Reserve, French Guiana (NOU-11). TLS data were acquired from a further single 0.25 ha censused plot in *Eucalyptus* spp. open forest in Karawatha Forest Park, Australia (KARA-001). The TLS data acquisition protocols employed across each of these four plots is defined in chapter 4 section 4.2. A description of the two-stage scan co-registration and noise reduction methods, used in the construction of the plot-level point clouds from these TLS data, can be found in chapter 4 section 4.1.

As described in chapter 5, tree-level point clouds were extracted from each of the four plot-level point clouds using the *treeseg* algorithm. For the work in this thesis, the objective was to extract all trees inside each plot with $D \geq 0.2$ m. Chapter 5 section 5.4 presents the results of the extraction, and illustrates the 441 tree-level point clouds that have been extracted from these TLS data. Of note, during the application of *treeseg*, each of the tree-level point clouds were downsampled via voxel grid downsampling using a voxel grid edge length of 0.04 m (see chapter 5 section 5.2.3).

Each of these 441 tree-level point clouds were then reconstructed using the shape fitting algorithm of Raumonen et al. (2013), as described in chapter 6 section 6.2.1, to construct 441 corresponding QSMs. Each of the QSMs were automatically optimised using the routine presented in chapter 6 section 6.2.3. That is, for each tree-level point cloud, the set, $\{d\}$, comprising the range of values of the input parameter, d, to be considered was defined as $\{d\} = \{0.040, 0.045, 0.050, ..., 0.140, 0.145, 0.150\}$. As dictated by the empirical calibration in chapter 6 section 6.3.1.1, the scalar parameters of the automatic QSM optimisation routine were defined as $S_{CV_V} = 5$ and $S_{T_c} = 0.95$. By using this automatic QSM optimisation routine, and holding N constant at 3, all further input parameters of the of Raumonen et al. (2013) algorithm were self-selecting. 20 QSMs were constructed per tree-level point cloud, for each value of d, resulting in the construction of a total of 460 QSMs. Finally, in line with the requirement for the most uniformly distributed tree-level point clouds possible as input to the reconstruction process, the downsampled point clouds derived from the *treeseg* algorithm were used as is. Therefore, to conform with the height-resolved mean nearest neighbour distance across these tree-level point clouds, all cylinders with a radius less than 0.025 m were pruned from the QSMs.

Tree-scale volume, $\bar{V}$, was derived as the mean tree-scale volume across the 20 QSMs that correspond to the optimised value of d, from which the standard deviation of the

volume, σ_V, was also calculated. From the estimate of $\bar{V}$, the TLS-derived estimate of tree-scale AGB, $\mathrm{AGB}_{\mathrm{TLS}}$, is defined as:

$$\mathrm{AGB}_{\mathrm{TLS}} = \bar{V}\rho \tag{7.1}$$

Where ρ is the plot-average wood density. The definition of ρ across each of the four considered plots can be found in chapter 4 section 4.2. Accompanying each estimate of $\mathrm{AGB}_{\mathrm{TLS}}$, is an approximation of the relative uncertainty of the estimate, U_t. This is defined utilising the 95 % confidence intervals that are derived from the observed spread in tree-scale volume across the 20 QSMs corresponding to the optimised value of d, defined as:

$$U_t = \frac{\sigma_V \; z_{0.025}}{\bar{V}} \tag{7.2}$$

Where $z_{0.025}$ is the standard score. As was demonstrated in chapter 6 section 6.3, this approximation of U_t reasonably captured the observed error in $\mathrm{AGB}_{\mathrm{TLS}}$ when applied to validation data; making it a reasonable proxy for the expected tree-level uncertainty in $\mathrm{AGB}_{\mathrm{TLS}}$.

7.2.2 Allometric estimates of tree- and plot-scale AGB

Allometric estimation of tree-scale AGB, $\mathrm{AGB}_{\mathrm{Allom}}$ is predicted by the NLS allometric model developed in chapter 2. This allometric model, employing a single amalgamated allometric regressor, takes the form:

$$\mathrm{AGB} = 0.179(D^2 H \rho)^{1.102} + 80.235 \tag{7.3}$$

For each tree, the regressor D is obtained from the tree-level point cloud via the RANSAC cylinder fit method presented in chapter 5 section 5.3.8. The regressor H is obtained from the tree-level point cloud via the Euclidean distance between the minimum and maximum z-coordinates through the cloud. The regressor ρ is the plot-average wood density, as defined for each of the four plots in chapter 4 section 4.2.

The relative uncertainty in the estimate of tree-scale AGB, U_t, is derived from the methods described in chapter 2. That is, the two additive contributors to U_t, the tree-

scale allometric model uncertainty and the tree-scale measurement uncertainty, are derived from non-linear quantile regression and Monte Carlo simulation respectively. The relative uncertainty in the estimate of plot-scale AGB, U_p, is also derived from the methods described in chapter 2. That is, the two additive contributors to U_p, the plot-scale allometric model uncertainty and the plot-scale measurement uncertainty, are derived from the wild bootstrap and Monte Carlo simulation respectively.

7.3 Results

The following four sections 7.3.1, 7.3.2, 7.3.3 and 7.3.4 present the results from the individual forest plots KARA-001, LPG-01, CAX-A and NOU-11 respectively. Each of these sections commence with an illustration of the optimised QSMs for each of the extracted tree-level point clouds. In each instance, the QSM shown is that of the 20 in the set corresponding to the optimised value of d, whose tree-scale volume lies closest to the mean volume. Following this, a table is provided with a number of metrics comparing the TLS- and allometric-derived estimates of tree- and plot-scale AGB. The first of these metrics is the tree-level percentage difference between $\mathrm{AGB_{TLS}}$ and $\mathrm{AGB_{Allom}}$, $\Delta\%_t$, defined as:

$$\Delta\%_t = \frac{|\mathrm{AGB_{TLS}} - \mathrm{AGB_{Allom}}|}{(\frac{\mathrm{AGB_{TLS}} + \mathrm{AGB_{Allom}}}{2})} \tag{7.4}$$

The mean tree-level percentage difference, $\overline{\Delta\%_t}$, across the N individuals inside each plot, is defined as:

$$\overline{\Delta\%_t} = \sum_{i=1}^{N} \Delta\%_{t,i} \tag{7.5}$$

The spread of $\Delta\%_t$, about $\overline{\Delta\%_t}$, is defined by the standard deviation of the mean tree-level percentage difference, $\sigma_{\overline{\Delta\%_t}}$, defined as:

$$\sigma_{\overline{\Delta\%_t}} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\Delta\%_t - \overline{\Delta\%_t})^2} \tag{7.6}$$

The TLS-derived plot-scale estimate of AGB, $\mathrm{AGB_{TLS},}_p$, and the allometric estimate of plot-scale AGB, $\mathrm{AGB_{Allom},}_p$, is defined as the summation of the tree-scale AGB across

the N individuals inside each plot, defined as:

$$\mathrm{AGB}_{\mathrm{TLS},p} = \sum_{i=1}^{N} \mathrm{AGB}_{\mathrm{TLS},i} \tag{7.7}$$

$$\mathrm{AGB}_{\mathrm{Allom},p} = \sum_{i=1}^{N} \mathrm{AGB}_{\mathrm{Allom},i} \tag{7.8}$$

Finally, the plot-level percentage difference between $\mathrm{AGB}_{\mathrm{TLS},p}$ and $\mathrm{AGB}_{\mathrm{Allom},p}$, $\Delta\%_p$, is defined as:

$$\Delta\%_p = \frac{|\mathrm{AGB}_{\mathrm{TLS},p} - \mathrm{AGB}_{\mathrm{Allom},p}|}{(\frac{\mathrm{AGB}_{\mathrm{TLS},p} + \mathrm{AGB}_{\mathrm{Allom},p}}{2})} \tag{7.9}$$

Following the definition of these metrics across the respective plots, each of the sections then presents a figure that directly compares $\mathrm{AGB}_{\mathrm{TLS}}$ with $\mathrm{AGB}_{\mathrm{Allom}}$. Each section then concludes by presenting four examples of the QSMs from across the plot overlaid the tree-level point cloud from which they were constructed.

Section 7.3.5 then presents a general overview of these results and provides a summary of the comparisons between the TLS- and allometric-derived estimates of AGB. Finally, section 7.3.6 presents a new allometric model that has been generated by supplementing the allometric dataset of Chave et al. (2014) with these new TLS data.

7.3.1 Karawatha Forest Park, Australia (KARA-001)

As described in chapter 4 section 4.2.1, TLS data were acquired from the 0.25 ha censused *Eucalyptus* spp. open forest plot KARA-001 in Karawatha Forest Park, Australia. As shown in chapter 5 section 5.4.1, 28 trees were extracted from these TLS data using the *treeseg* algorithm. AGB_{TLS} has been estimated for the 28 trees via application of the Raumonen et al. (2013) shape fitting algorithm and the automatic QSM optimisation routine. Figure 7.1 illustrates the optimised QSM constructed from each of the tree-level point clouds. AGB_{Allom} has been estimated for the 28 trees through the NLS pan-tropical allometric model. Despite this plot not being tropical forest, the application of this model is suitable here as the allometric dataset of Chave et al. (2014) contains 46 harvested *Eucalyptus* spp. individuals from open forest in Australia.

Table 7.1 compares these TLS- and allometric-derived estimates of tree- and plot-scale AGB, and their uncertainties. It can be seen that $AGB_{p,TLS}$ is 34.5 % larger than $AGB_{Allom,p}$. It can also be seen that mean uncertainty in the estimate of AGB_{Allom} is 85.8 % larger than the mean uncertainty in the estimate of AGB_{TLS}.

$AGB_{p,TLS}$	$\bar{U}_{TLS,t}$	$AGB_{Allom,p}$	$\bar{U}_{Allom,t}$	$\Delta\%_p$	$\overline{\Delta\%_t}$	$\sigma_{\Delta\%_t}$
50 338 kg	19.2 %	35 531 kg	105.0 %	34.5 %	29.0 %	12.9 %

Table 7.1: Metrics comparing the TLS- and allometric-derived estimates of tree- and plot-scale AGB, and uncertainty, for the 28 individuals across KARA-001.

Figure 7.2 presents a direct comparison of AGB_{TLS} and AGB_{Allom}. There is a strong trend that as tree-scale AGB increases, $\Delta\%_t$ tends towards a larger value (i.e., AGB_{TLS} is increasingly larger than AGB_{Allom}). It is worth noting that this is near-identical behaviour to that observed in chapter 6 section 6.3.1.2 across the considered validation data from Rushworth State Forest, Australia.

Finally, figure 7.3 presents four examples of QSMs from KARA-001 overlaying the tree-level point cloud from which they were constructed. Figure 7.3a presents the QSM producing the largest value of $\Delta\%_t$. Figure 7.3b presents the QSM producing the value of $\Delta\%_t$ closest to $\overline{\Delta\%_t}$. Figure 7.3c presents the QSM producing the smallest value of $\Delta\%_t$. Figure 7.3d then presents the QSM contributing the largest value of AGB_{TLS} to $AGB_{p,TLS}$. In each instance it appears that the QSM has faithfully reconstructed the woody structure of the tree. However, it can be seen in figure 7.3b that the arrangement of the cylinders through the stem appears somewhat inconsistent.

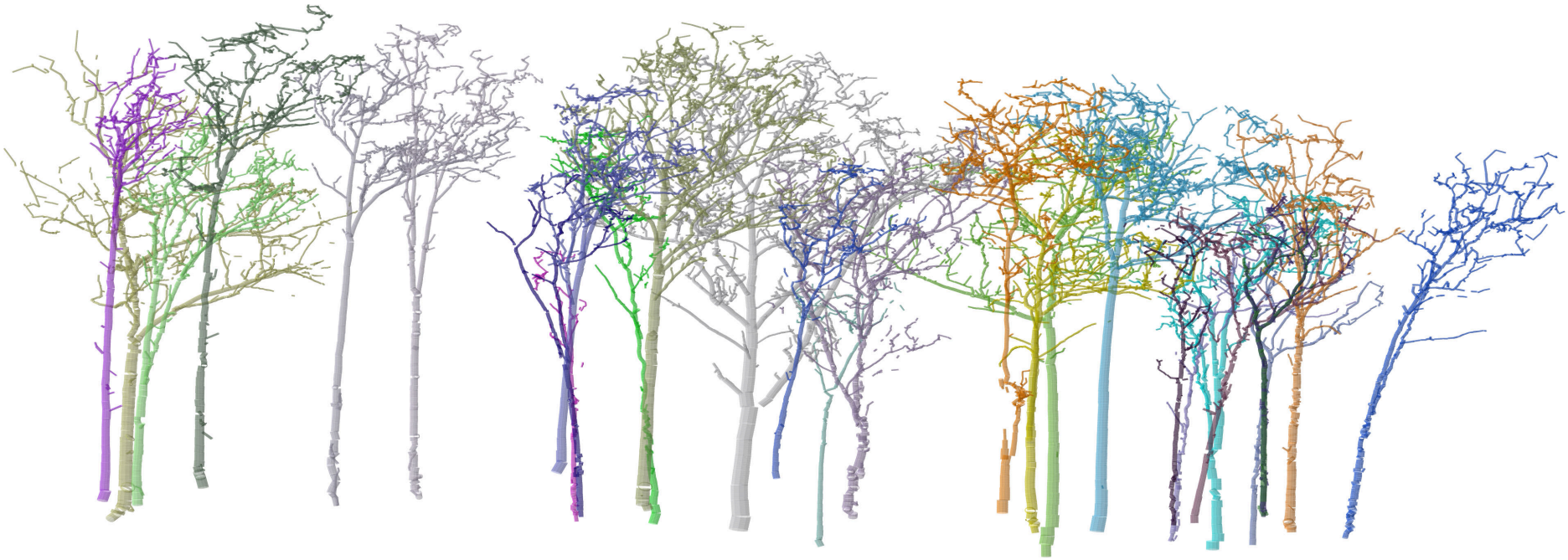


Figure 7.1: The optimised QSMs constructed from the 28 extracted tree-level point clouds across KARA-001.

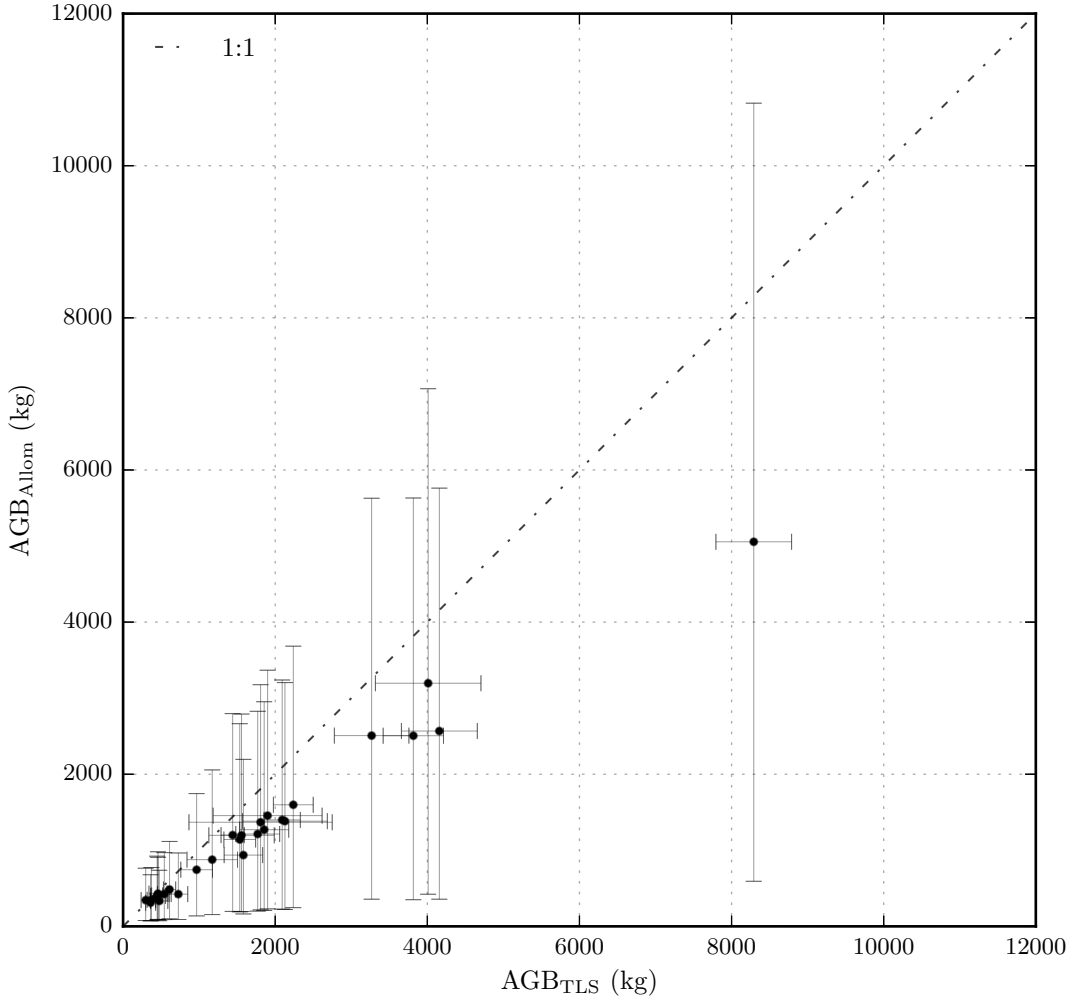
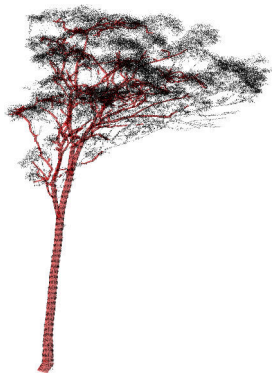
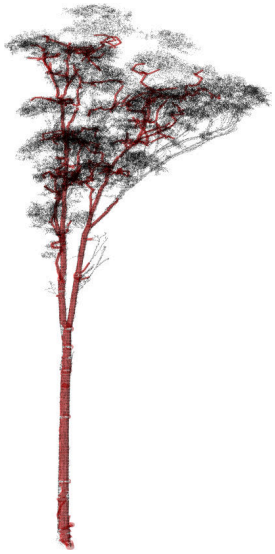


Figure 7.2: Comparison between the TLS- and allometric-derived estimates of tree-scale AGB for the 28 individuals across the plot KARA-001.



(a) AGB_{TLS} = 1582 kg, AGB_{Allom} = 937 kg

(b) AGB_{TLS} = 1172 kg, AGB_{Allom} = 877 kg



(c) AGB_{TLS} = 372 kg, AGB_{Allom} = 348 kg

(d) AGB_{TLS} = 8291 kg, AGB_{Allom} = 5055 kg

Figure 7.3: Examples of four QSMs (red) and the underlying tree-level point clouds (black) from which they were constructed from across the plot KARA-001; the QSM in: (a) produced the largest $\Delta\%_t$, (b) produced the $\Delta\%_t$ closest to $\overline{\Delta\%_t}$, (c) produced the smallest $\Delta\%_t$ and (d) was the largest individual by AGB that was observed across the plot.

7.3.2 Lopé National Park, Gabon (LPG-01)

As described in chapter 4 section 4.2.2, TLS data were acquired from the 1 ha censused moist, Terra Firme, lowland, mixed species, old-growth tropical forest plot LPG-01 in Lopé National Park, Gabon. As shown in chapter 5 section 5.4.2, 107 trees were extracted from these TLS data using the *treeseg* algorithm. $\mathrm{AGB}_{\mathrm{TLS}}$ has been estimated for the 107 trees via application of the Raumonen et al. (2013) shape fitting algorithm and the automatic QSM optimisation routine. Figure 7.4 illustrates the optimised QSM constructed from each of the tree-level point clouds. $\mathrm{AGB}_{\mathrm{Allom}}$ has been estimated for the 107 through the NLS pan-tropical allometric model.

Table 7.2 compares these TLS- and allometric-derived estimates of tree- and plot-scale AGB, and their uncertainties. It can be seen that $\mathrm{AGB}_{p,\mathrm{TLS}}$ is 16.4 % larger than $\mathrm{AGB}_{\mathrm{Allom},p}$. It can also be seen that mean uncertainty in the estimate of $\mathrm{AGB}_{\mathrm{Allom}}$ is 82.8 % larger than the mean uncertainty in the estimate of $\mathrm{AGB}_{\mathrm{TLS}}$.

$\mathrm{AGB}_{p,\mathrm{TLS}}$	$\bar{U}_{\mathrm{TLS},t}$	$\mathrm{AGB}_{\mathrm{Allom},p}$	$\bar{U}_{\mathrm{Allom},t}$	$\Delta\%_p$	$\overline{\Delta\%_t}$	$\sigma_{\Delta\%_t}$
655 884 kg	20.0 %	556 195 kg	102.8 %	16.4 %	49.7 %	34.5 %

Table 7.2: Metrics comparing the TLS- and allometric-derived estimates of tree- and plot-scale AGB, and uncertainty, for the 107 individuals across LPG-01.

Figure 7.5 presents a direct comparison of $\mathrm{AGB}_{\mathrm{TLS}}$ and $\mathrm{AGB}_{\mathrm{Allom}}$ across a series of four AGB ranges. Figure 7.6 then presents the overall comparison of $\mathrm{AGB}_{\mathrm{TLS}}$ and $\mathrm{AGB}_{\mathrm{Allom}}$. It can be seen that across the AGB range from 200–5000 kg, the value of $\mathrm{AGB}_{\mathrm{TLS}}$ is almost always larger than $\mathrm{AGB}_{\mathrm{Allom}}$. Above this range, the sign of $\Delta\%_t$ is more evenly spread. That is, of 8 individuals above this AGB range, across half, $\mathrm{AGB}_{\mathrm{TLS}}$ is larger than $\mathrm{AGB}_{\mathrm{Allom}}$, whilst the opposite is true of the remaining half. Notably, the largest tree that is observed across all the data considered in this thesis, with $D = 2.52$ m and $H = 43.4$ m, produced values of $\mathrm{AGB}_{\mathrm{TLS}} = 105\,692$ kg and $\mathrm{AGB}_{\mathrm{Allom}} = 127\,765$ kg.

Finally, figure 7.7 presents four examples of QSMs from LPG-01 overlaying the tree-level point cloud from which they were constructed. Figure 7.7a presents the QSM producing the largest value of $\Delta\%_t$. Figure 7.7b presents the QSM producing the value of $\Delta\%_t$ closest to $\overline{\Delta\%_t}$. Figure 7.7c presents the QSM producing the smallest value of $\Delta\%_t$. Figure 7.7d then presents the QSM contributing the largest value of $\mathrm{AGB}_{\mathrm{TLS}}$ to $\mathrm{AGB}_{p,\mathrm{TLS}}$. In each instance, it appears visually that the QSM has faithfully reconstructed the woody structure of the tree. The most apparent discrepancy can be seen

in figure 7.7c, where the individual has a complex above-surface rooting system that is insufficiently represented through a single cylinder.

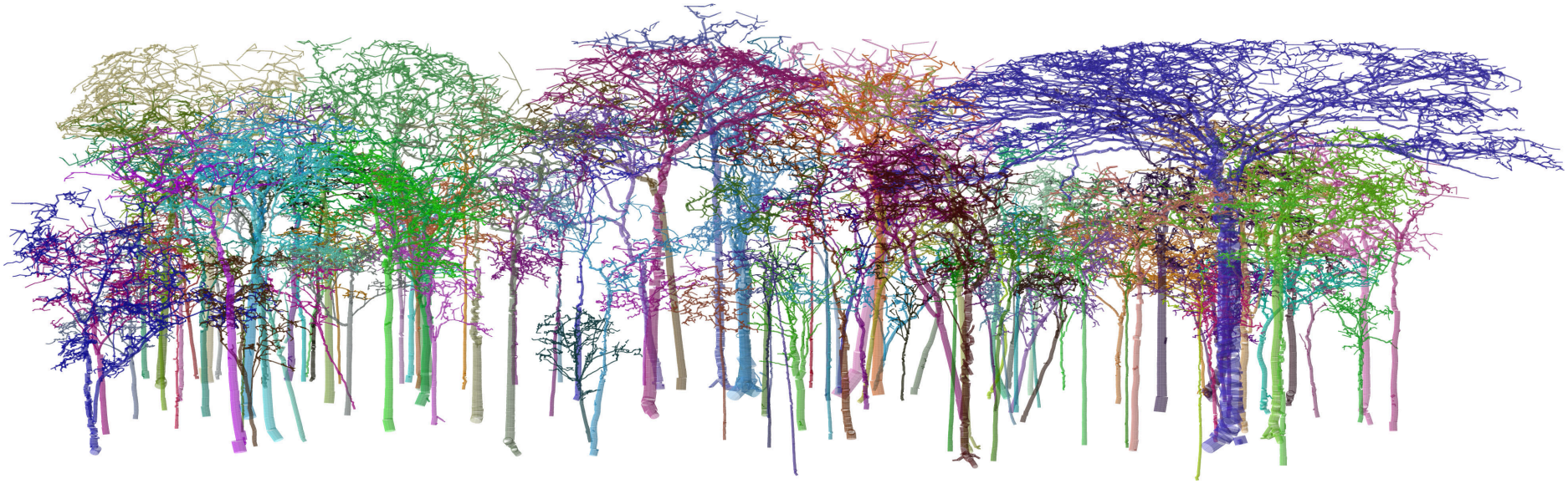


Figure 7.4: The optimised QSMs constructed from the 107 extracted tree-level point clouds across LPG-01.

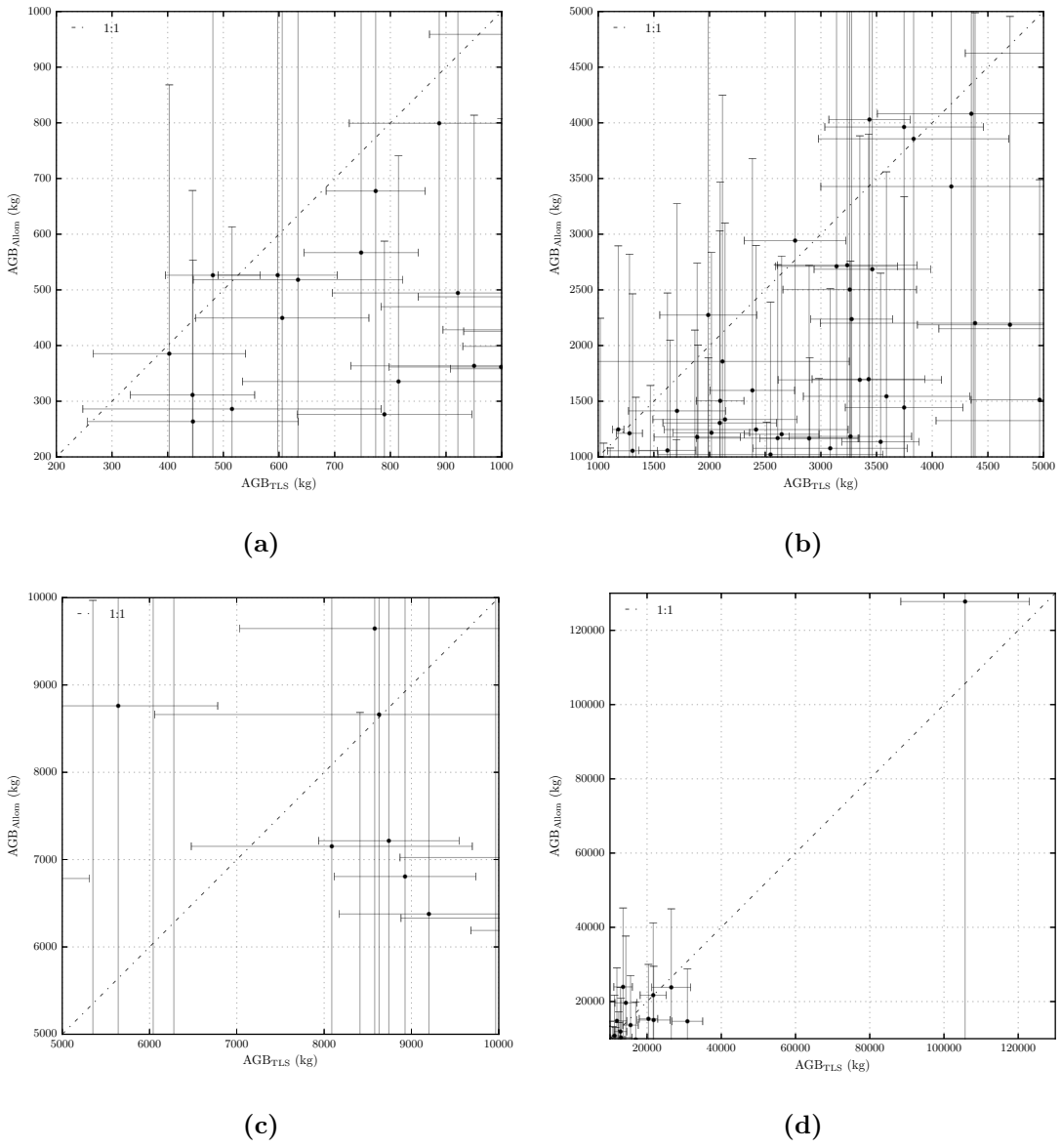


Figure 7.5: Comparison between the TLS- and allometric-derived estimates of tree-scale AGB across the plot LPG-01; for the AGB ranges: (a) 200–1000 kg, (b) 1000–5000 kg, (c) 5000–10 000 kg and (d) 10 000–125 000 kg.

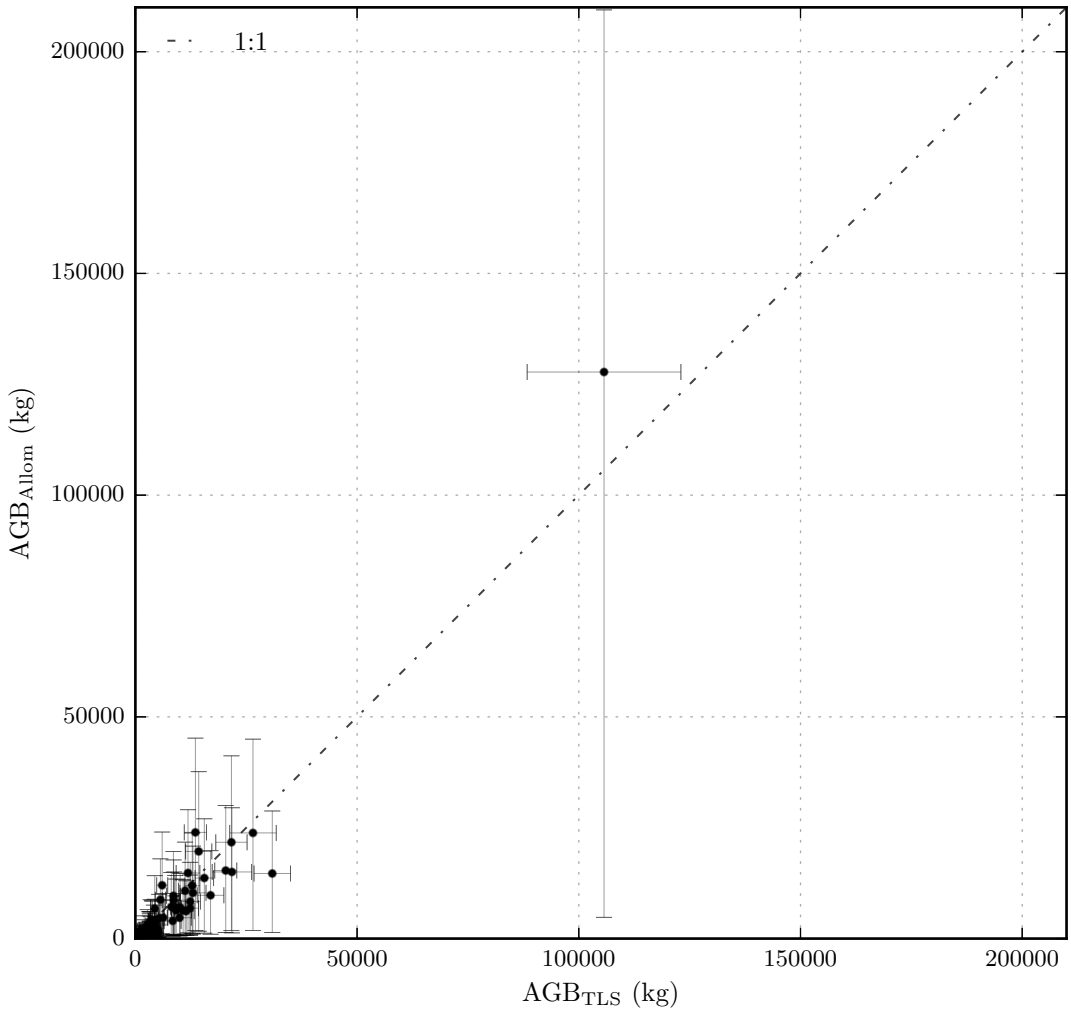
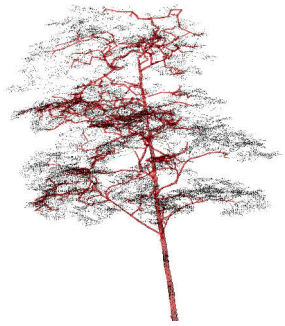
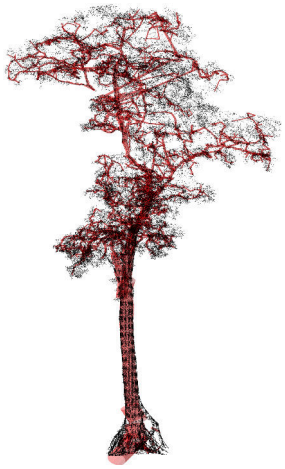


Figure 7.6: Overall comparison between the TLS- and allometric-derived estimates of tree-scale AGB for the 107 individuals across the plot LPG-01.

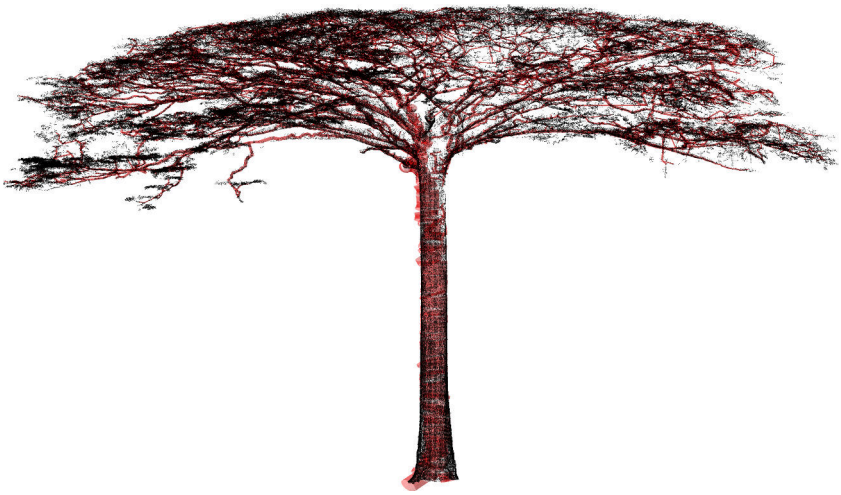


(a) $\mathrm{AGB}_{\mathrm{TLS}} = 2983\,\mathrm{kg}$, $\mathrm{AGB}_{\mathrm{Allom}} = 727\,\mathrm{kg}$

(b) $\mathrm{AGB}_{\mathrm{TLS}} = 5641\,\mathrm{kg}$, $\mathrm{AGB}_{\mathrm{Allom}} = 8761\,\mathrm{kg}$



(c) $\mathrm{AGB}_{\mathrm{TLS}} = 21\,660\,\mathrm{kg}$, $\mathrm{AGB}_{\mathrm{Allom}} = 21\,719\,\mathrm{kg}$



(d) $\mathrm{AGB}_{\mathrm{TLS}} = 105\,692\,\mathrm{kg}$, $\mathrm{AGB}_{\mathrm{Allom}} = 127\,765\,\mathrm{kg}$

Figure 7.7: Examples of four QSMs (red) and the underlying tree-level point clouds (black) from which they were constructed from across the plot LPG-01; the QSM in: (a) produced the largest $\Delta\%_t$, (b) produced the $\Delta\%_t$ closest to $\overline{\Delta\%_t}$, (c) produced the smallest $\Delta\%_t$ and (d) was the largest individual by AGB that was observed across the plot.

7.3.3 Caixuanã National Forest, Brazil (CAX-A)

As described in chapter 4 section 4.2.3, TLS data were acquired from the 1 ha censused moist, Terra Firme, lowland, mixed species, old-growth tropical forest plot CAX-A in Caixuanã National Forest, Brazil. As shown in chapter 5 section 5.4.3, 151 trees were extracted from these TLS data using the *treeseg* algorithm. AGB_{TLS} has been estimated for the 151 trees via application of the Raumonen et al. (2013) shape fitting algorithm and the automatic QSM optimisation routine. Figure 7.8 illustrates the optimised QSM constructed from each of the tree-level point clouds. AGB_{Allom} has been estimated for the 151 trees through the NLS pan-tropical allometric model.

Table 7.3 compares these TLS- and allometric-derived estimates of tree- and plot-scale AGB, and their uncertainties. It can be seen that $AGB_{p,TLS}$ is 5.3 % larger than $AGB_{Allom,p}$. It can also be seen that mean uncertainty in the estimate of AGB_{Allom} is 85.9 % larger than the mean uncertainty in the estimate of AGB_{TLS}.

$AGB_{p,TLS}$	$\bar{U}_{TLS,t}$	$AGB_{Allom,p}$	$\bar{U}_{Allom,t}$	$\Delta\%_p$	$\overline{\Delta\%_t}$	$\sigma_{\Delta\%_t}$
434 448 kg	18.9 %	412 167 kg	104.8 %	5.3 %	33.4 %	23.6 %

Table 7.3: Metrics comparing the TLS- and allometric-derived estimates of tree- and plot-scale AGB, and uncertainty, for the 151 individuals across CAX-A.

Figure 7.9 presents a direct comparison of AGB_{TLS} and AGB_{Allom} across a series of four AGB ranges. Figure 7.10 then presents the overall comparison of AGB_{TLS} and AGB_{Allom}. It can be seen that across the AGB range from 200–1000 kg, similar to plot LPG-01, the value of AGB_{TLS} is nearly always larger than AGB_{Allom}. However, across the AGB range from 1000–5000 kg this tendency is less pronounced, explaining the overall reduction in $\Delta\%_p$ in comparison to LPG-01. This overall reduction in $\Delta\%_p$ is also attributed to corroborating values of AGB_{TLS} and AGB_{Allom} encountered across most of the larger individuals.

Finally, figure 7.11 presents four examples of QSMs from CAX-A overlaying the tree-level point cloud from which they were constructed. Figure 7.11a presents the QSM producing the largest value of $\Delta\%_t$. Figure 7.11b presents the QSM producing the value of $\Delta\%_t$ closest to $\overline{\Delta\%_t}$. Figure 7.11c presents the QSM producing the smallest value of $\Delta\%_t$. Figure 7.11d then presents the QSM contributing the largest value of AGB_{TLS} to $AGB_{p,TLS}$. In each instance it appears that the QSM has faithfully reconstructed the woody structure of the tree. Most notable, in figure 7.11d, is the appearance of a poorly

modelled first order branch at the commencement of the crown. It can been seen that this section of the tree-level point cloud appears poorly sampled, which is probably the underlying cause.

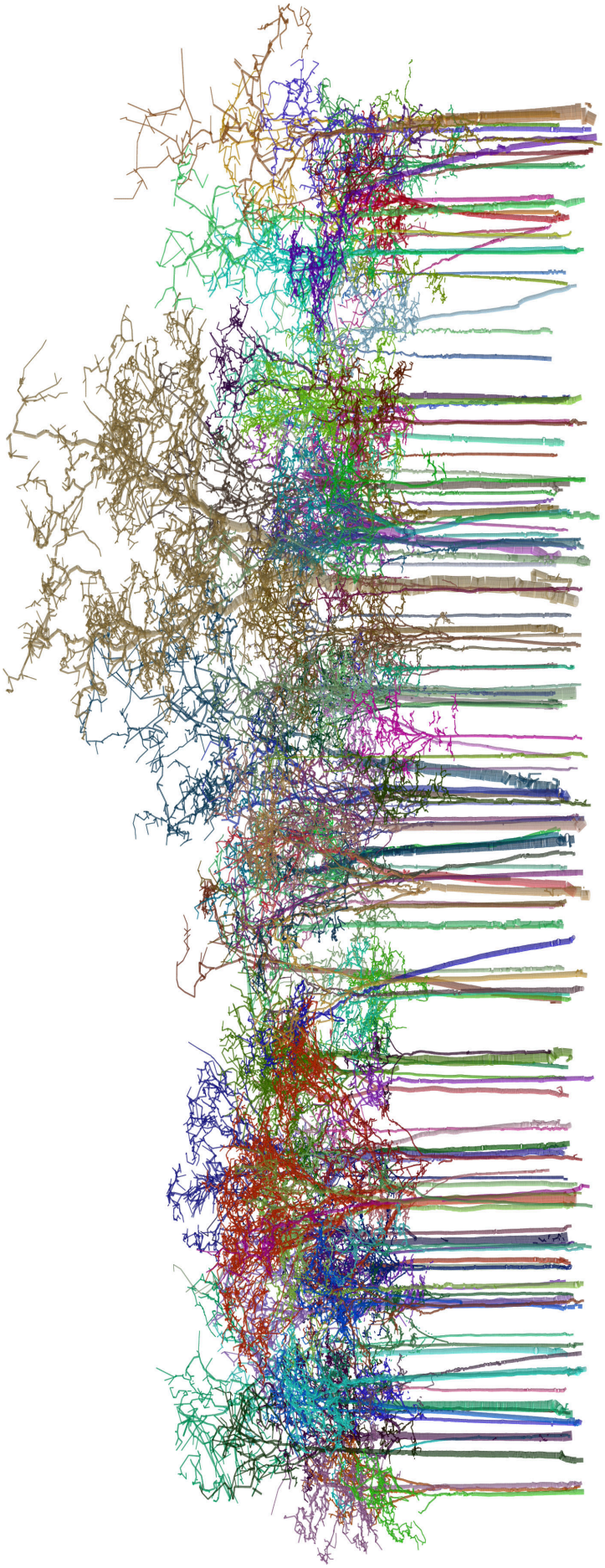


Figure 7.8: The optimised QSMs constructed from the 151 extracted tree-level point clouds across CAX-A.

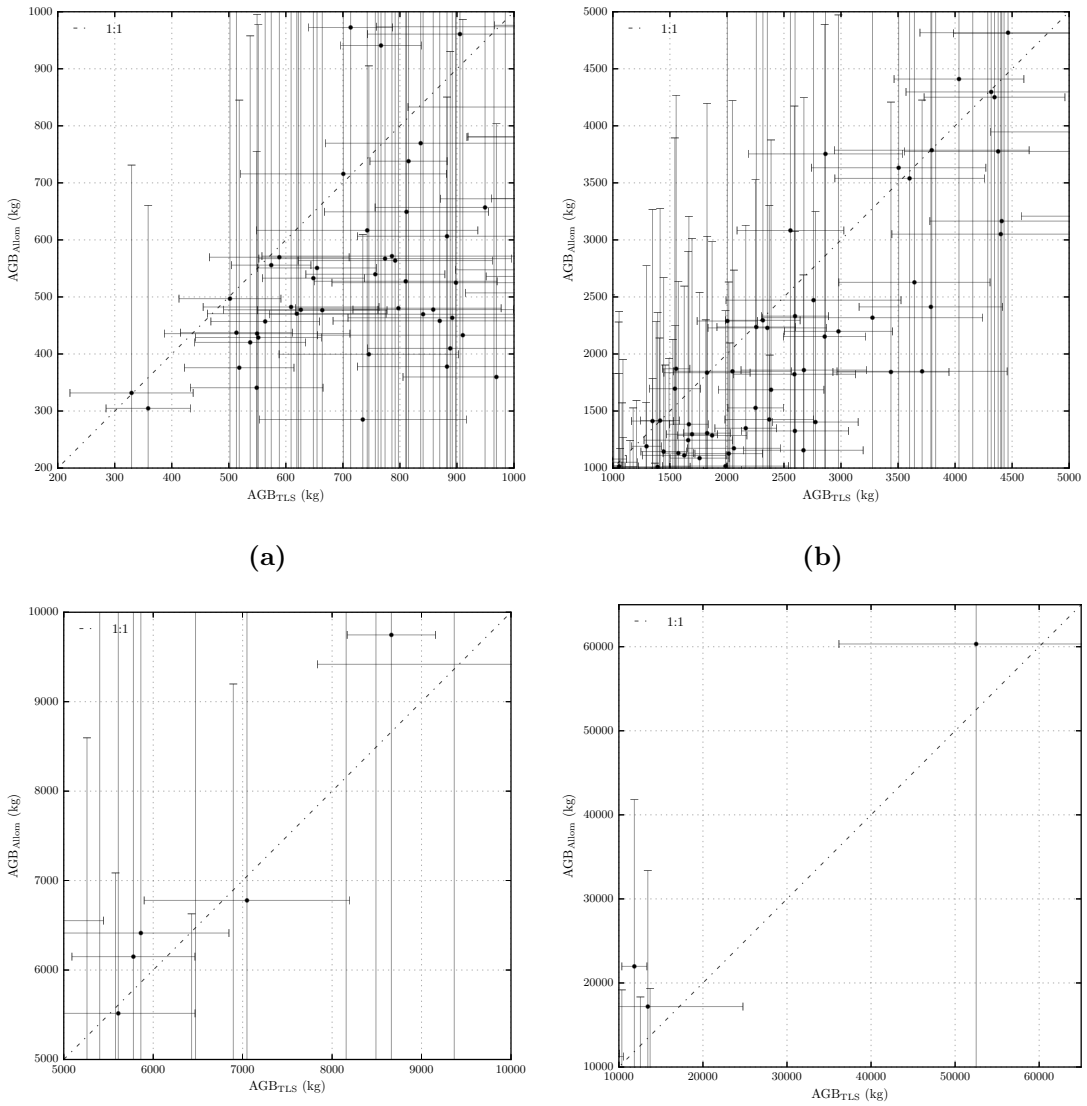


Figure 7.9: Comparison between the TLS- and allometric-derived estimates of tree-scale AGB across the plot CAX-A; for the AGB ranges: (a) 200–1000 kg, (b) 1000–5000 kg, (c) 5000–10 000 kg and (d) 10 000–125 000 kg.

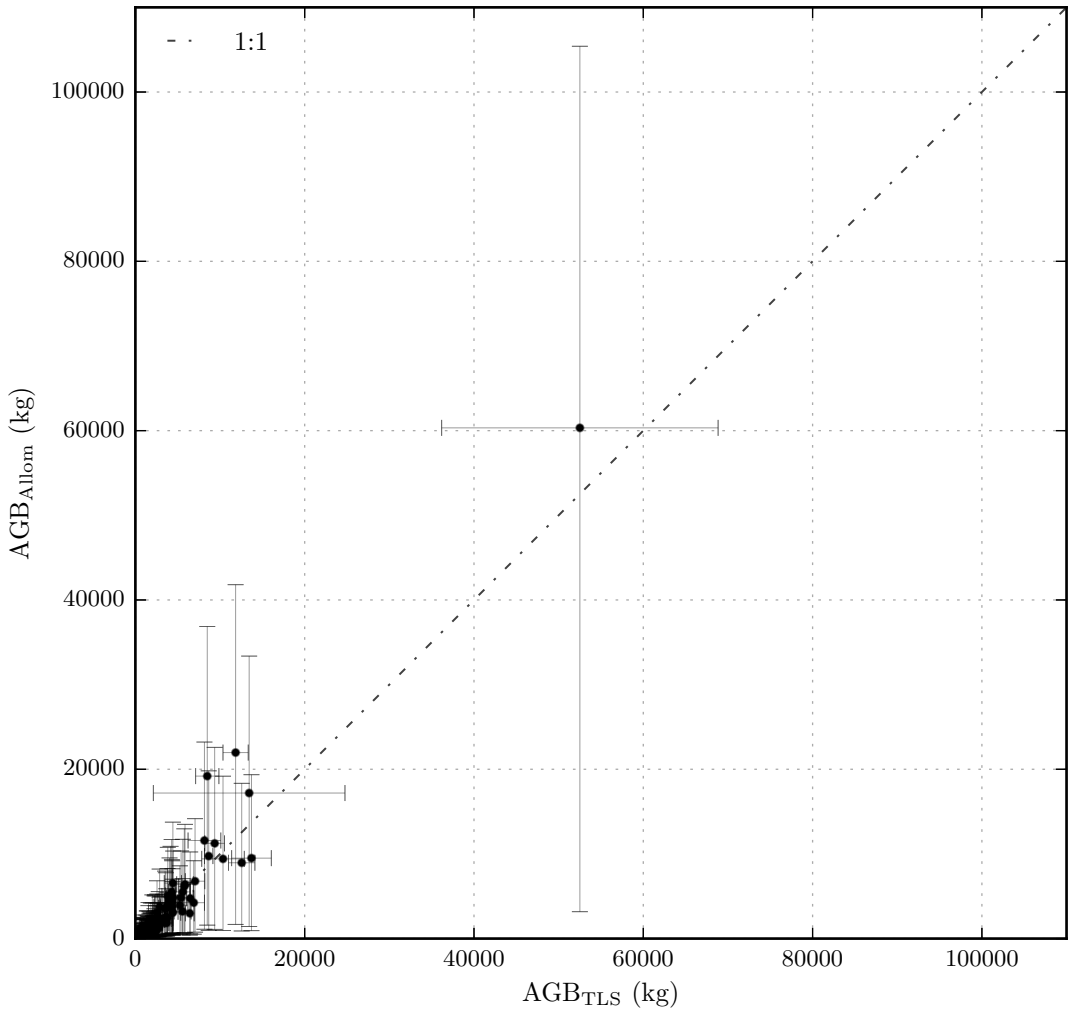
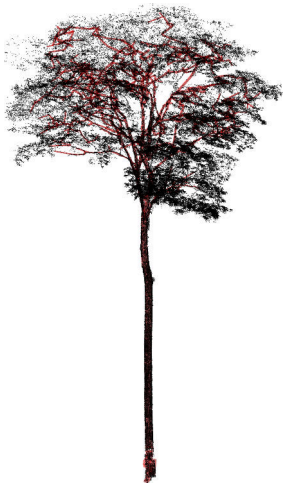
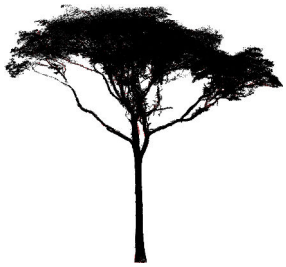


Figure 7.10: Overall comparison between the TLS- and allometric-derived estimates of tree-scale AGB for the 151 individuals across the plot CAX-A.



(a) AGB_{TLS} = 2671 kg, AGB_{Allom} = 1156 kg



(b) AGB_{TLS} = 13 714 kg, AGB_{Allom} = 9502 kg

(c) AGB_{TLS} = 4346 kg, AGB_{Allom} = 4251 kg

(d) AGB_{TLS} = 52 509 kg, AGB_{Allom} = 60 329 kg

Figure 7.11: Examples of four QSMs (red) and the underlying tree-level point clouds (black) from which they were constructed from across the plot CAX-A; the QSM in: (a) produced the largest $\Delta\%$, (b) produced the $\Delta\%$ closest to $\overline{\Delta\%}$, (c) produced the smallest $\Delta\%$ and (d) was the largest individual by AGB that was observed across the plot.

7.3.4 Nouragues Nature Reserve, French Guiana (NOU-11)

As described in chapter 4 section 4.2.4, TLS data were acquired from the 1 ha censused moist, Terra Firme, lowland, mixed species, old-growth tropical forest plot NOU-11 in Nouragues Nature Reserve, French Guiana. As shown in chapter 5 section 5.4.4, 155 trees were extracted from these TLS data using the *treeseg* algorithm. AGB_{TLS} has been estimated for the 155 trees via application of the Raumonen et al. (2013) shape fitting algorithm and the automatic QSM optimisation routine. Figure 7.12 illustrates the optimised QSM constructed from each of the tree-level point clouds. AGB_{Allom} has been estimated for the 155 trees through the NLS pan-tropical allometric model.

Table 7.4 compares these TLS- and allometric-derived estimates of tree- and plot-scale AGB, and their uncertainties. It can be seen that $AGB_{p,TLS}$ is 28.5 % larger than $AGB_{Allom,p}$. It can also be seen that mean uncertainty in the estimate of AGB_{Allom} is 90.6 % larger than the mean uncertainty in the estimate of AGB_{TLS}.

$AGB_{p,TLS}$	$\bar{U}_{TLS,t}$	$AGB_{Allom,p}$	$\bar{U}_{Allom,t}$	$\Delta\%_p$	$\overline{\Delta\%_t}$	$\sigma_{\Delta\%_t}$
599 681 kg	13.8 %	450 174 kg	104.4 %	28.5 %	47.0 %	27.6 %

Table 7.4: Metrics comparing the TLS- and allometric-derived estimates of tree- and plot-scale AGB, and uncertainty, for the 155 individuals across NOU-11.

Figure 7.13 presents a direct comparison of AGB_{TLS} and AGB_{Allom} across a series of four AGB ranges. Figure 7.14 then presents the overall comparison of AGB_{TLS} and AGB_{Allom}. In comparison to both LPG-01 and CAX-A, it can be seen that the value of AGB_{TLS} is near-universally larger than AGB_{Allom} across all 151 stems, invariant of AGB.

Finally, figure 7.15 presents four examples of QSMs from CAX-A overlaying the tree-level point cloud from which they were constructed. Figure 7.15a presents the QSM producing the largest value of $\Delta\%_t$. Figure 7.15b presents the QSM producing the value of $\Delta\%_t$ closest to $\overline{\Delta\%_t}$. Figure 7.15c presents the QSM producing the smallest value of $\Delta\%_t$. Figure 7.15d then presents the QSM contributing the largest value of AGB_{TLS} to $AGB_{p,TLS}$. As with the previous three plots, in each instance it appears that the QSM has faithfully reconstructed the woody structure of the tree. The most significant discrepancy is seen in figure 7.15a, where there is a large buttress formation, that having been modelled by a proportionally small, single cylinder, will have probably led to an underestimate in $\bar{V}$.

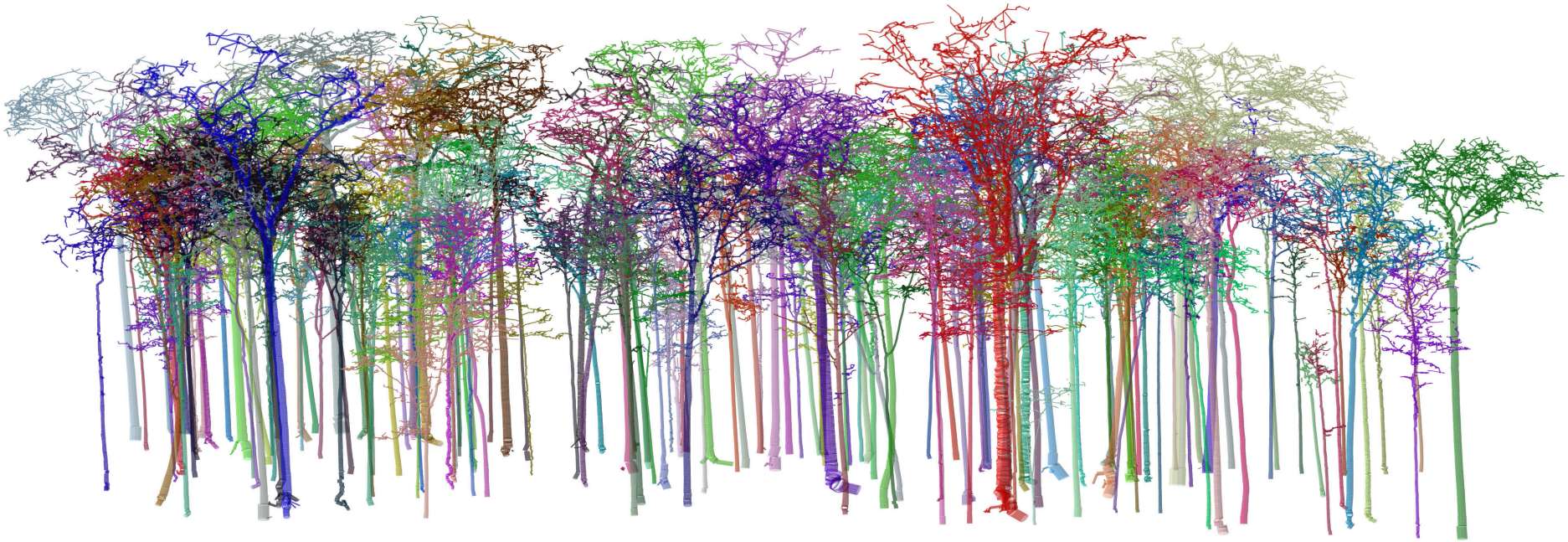


Figure 7.12: The optimised QSMs constructed from the 155 extracted tree-level point clouds across NOU-11.

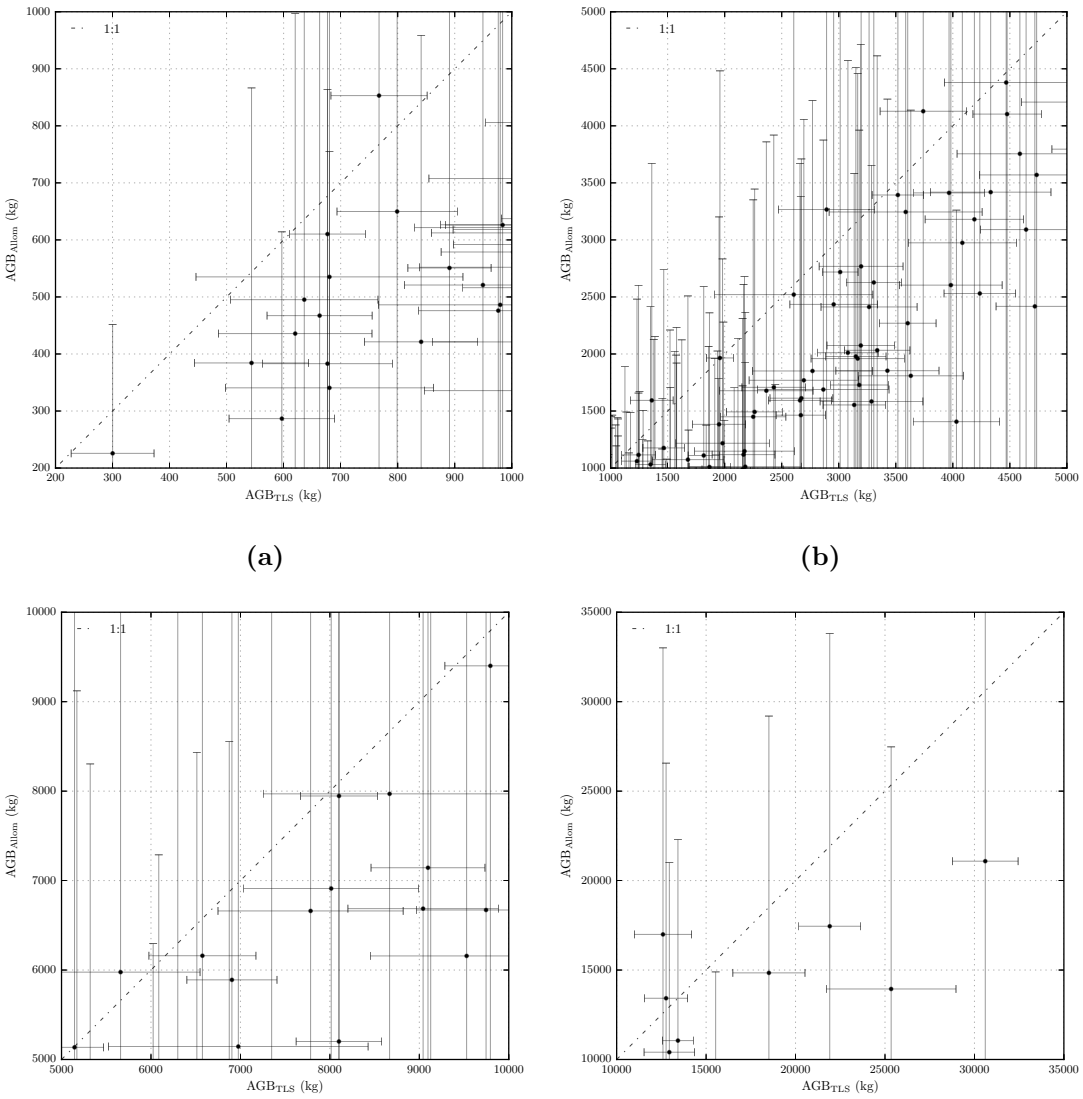


Figure 7.13: Comparison between the TLS- and allometric-derived estimates of tree-scale AGB across the plot NOU-11; for the AGB ranges: (a) 200–1000 kg, (b) 1000–5000 kg, (c) 5000–10 000 kg and (d) 10 000–125 000 kg.

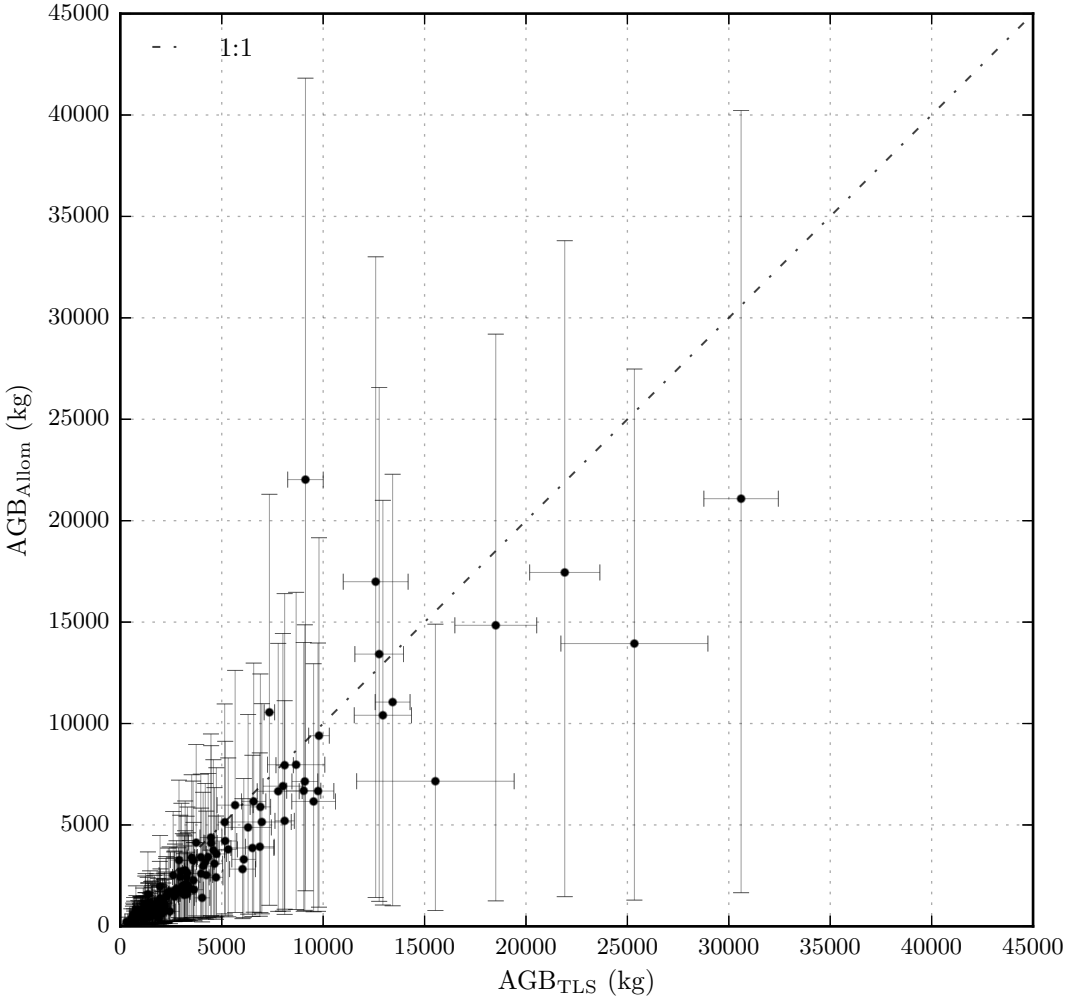
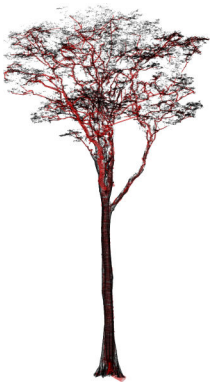


Figure 7.14: Overall comparison between the TLS- and allometric-derived estimates of tree-scale AGB for the 155 individuals across the plot NOU-11.



(a) AGB_{TLS} = 25 344 kg, AGB_{Allom} = 13 942 kg

(b) AGB_{TLS} = 6977 kg, AGB_{Allom} = 5145 kg

(c) AGB_{TLS} = 4476 kg, AGB_{Allom} = 4103 kg

(d) AGB_{TLS} = 30 606 kg, AGB_{Allom} = 21 085 kg

Figure 7.15: Examples of four QSMs (red) and the underlying tree-level point clouds (black) from which they were constructed from across the plot NOU-11; the QSM in: (a) produced the largest $\Delta\%$, (b) produced the $\Delta\%$ closest to $\overline{\Delta\%}$, (c) produced the smallest $\Delta\%$ and (d) was the largest individual by AGB that was observed across the plot.

7.3.5 Overview of the TLS and allometric estimates of AGB

This section combines the results presented in the previous sections for 441 individuals from across the forest plots KARA-001, LPG-01, CAX-A and NOU-11 to provide a brief overview. Table 7.5 compares the total TLS- and allometric-derived estimates of tree- and plot-scale AGB, and their uncertainties, across the four plots.

Plot code	AGB_{TLS} (kg)	AGB_{Allom} (kg)	$\Delta\%_p$	$\overline{\Delta\%_t}$	$\sigma_{\Delta\%_t}$
KARA-001	50 338	35 531	34.5	29.0	12.9
LPG-01	655 884	556 195	16.4	49.7	34.5
CAX-A	434 448	412 167	5.3	33.4	23.6
NOU-11	599 681	450 174	28.5	47.0	27.6
Total	1 740 351	1 454 067	17.9	41.8	28.6

Table 7.5: Metrics comparing the TLS- and allometric-derived estimates of plot-scale AGB, and uncertainty, for the 441 individuals across KARA-001, LPG-01, CAX-A and NOU-11.

Figure 7.16 presents a direct comparison of AGB_{TLS} and AGB_{Allom} from the four plots across a series of four AGB ranges. Figure 7.17 then presents the overall comparison of AGB_{TLS} and AGB_{Allom} from the four plots. In total, it can be seen that TLS-derived estimates of AGB come in at 1 740 351 kg, whilst the corresponding allometric derived estimate is 1 454 067 kg. This represents an overall percentage difference between the two methods of 17.9 %. Across the four plots, $\Delta\%$ ranges from 5.3–34.5 %. At the individual tree-level, $\overline{\Delta\%}$ is 41.8 %.

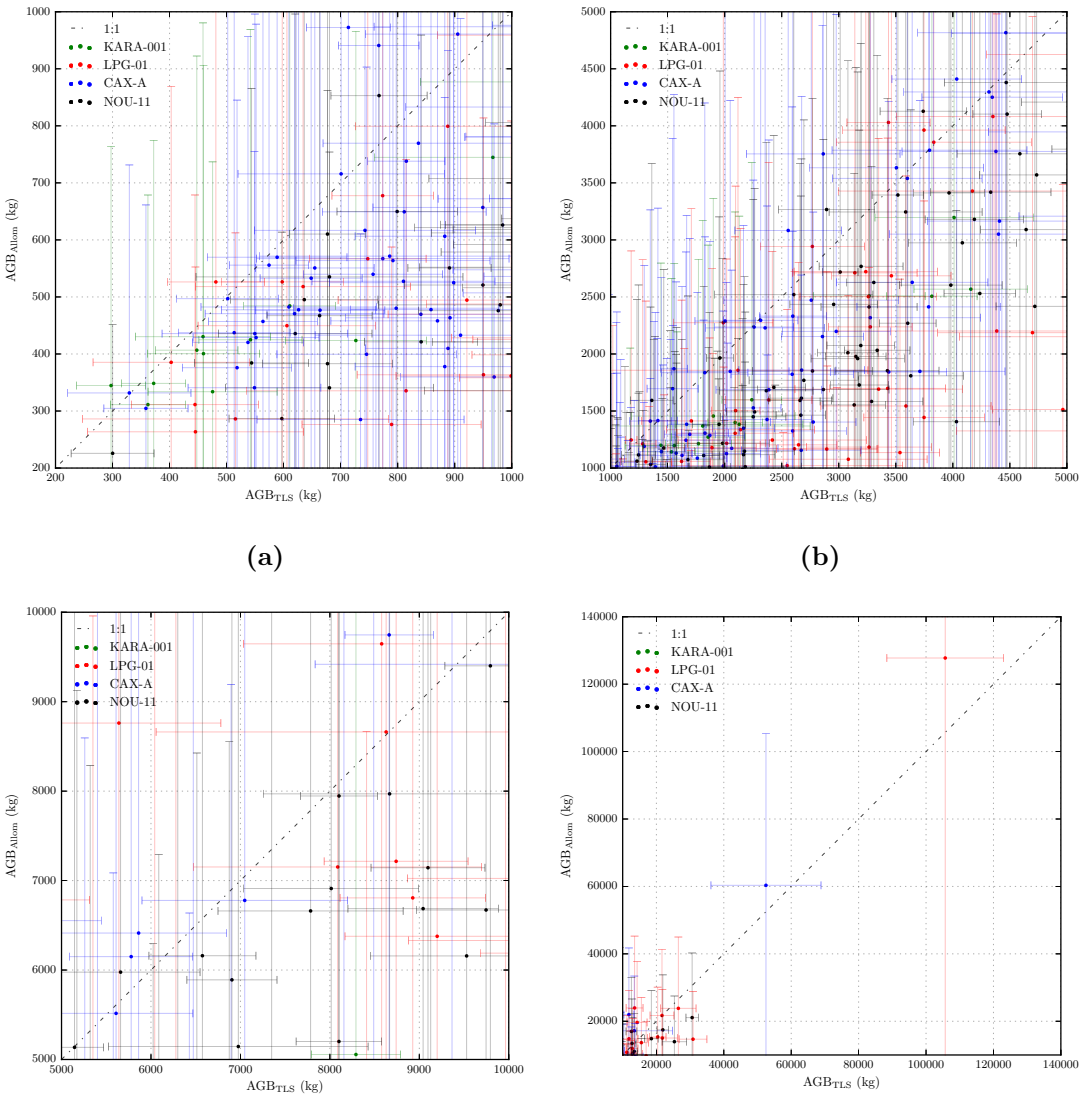


Figure 7.16: Comparison between the TLS- and allometric-derived estimates of tree-scale AGB across the plots KARA-001, LPG-01, CAX-A and NOU-11; for the AGB ranges: (a) 200–1000 kg, (b) 1000–5000 kg, (c) 5000–10 000 kg and (d) 10 000–125 000 kg.

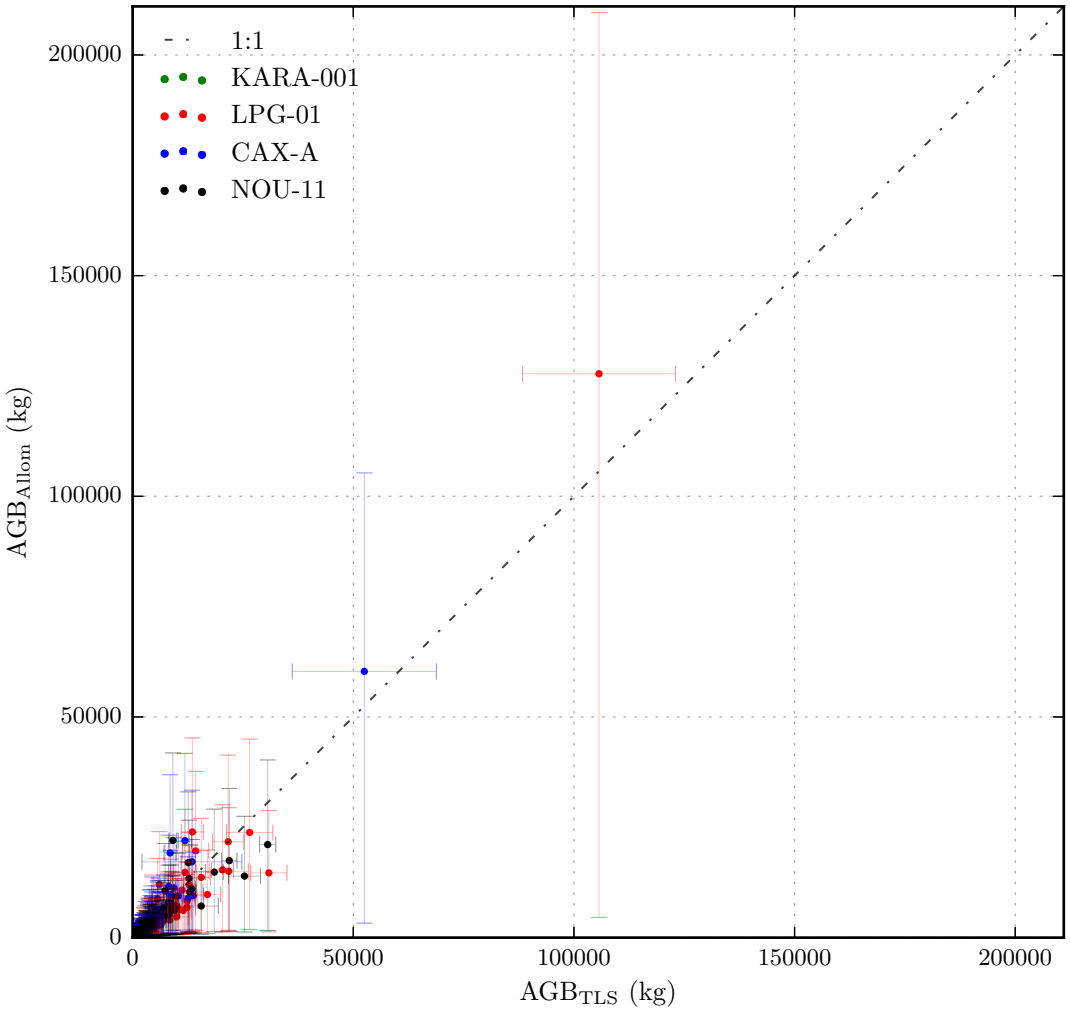


Figure 7.17: Overall comparison between the TLS- and allometric-derived estimates of tree-scale AGB for the 441 individuals across the plots KARA-001, LPG-01, CAX-A and NOU-11.

7.3.6 A new pan-tropical allometric model

This section presents the results from the construction of a new pan-tropical allometric model, that has been created by supplementing the allometric dataset of Chave et al. (2014), comprising the destructive harvest data of 4004 stems, with the 441 TLS-derived individuals presented in the previous sections. That is, the allometric dataset is augmented with these 441 TLS-derived estimates of AGB_{TLS}, D, H and the plot average wood density, ρ.

Figure 7.18 presents the distribution of stems, by D, with $D \geq 0.2\,\text{m}$, across the Chave et al. (2014) allometric dataset. Overlaid in this figure is the contribution to the augmented allometric dataset by the TLS data. It can be seen, that in total, this supplementation has increased the number of individuals with $D \geq 0.2\,\text{m}$, from 1529 to 1970 This results in the composition of individuals with $D \geq 0.2\,\text{m}$ across the allometric dataset rising from 38.2 % to 44.3 %.

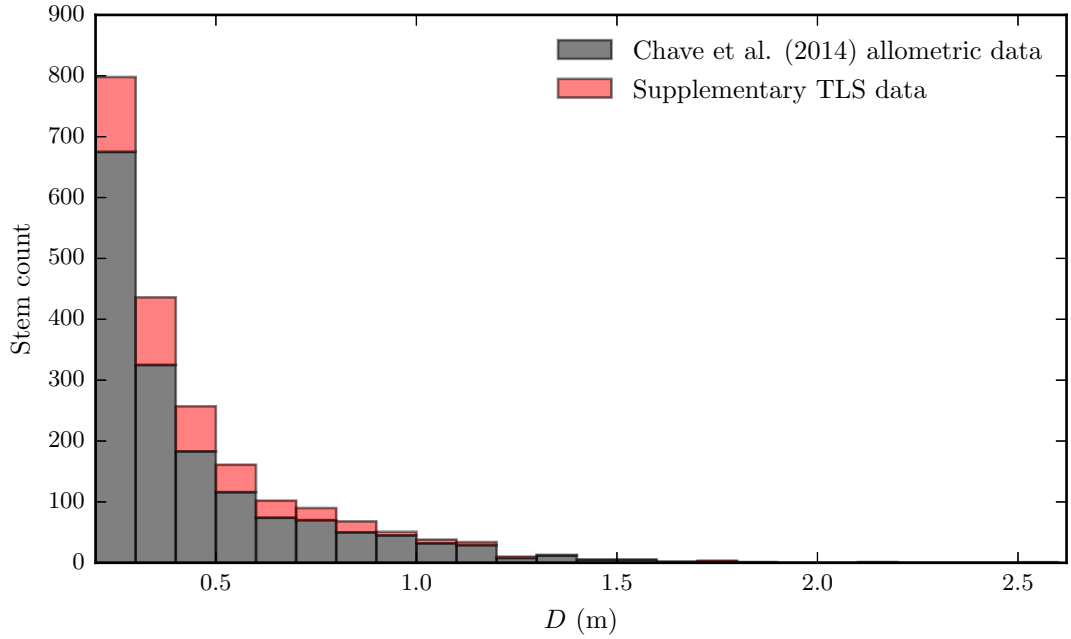


Figure 7.18: The distribution of stems, by D, with $D \geq 0.2\,\text{m}$ across the Chave et al. (2014) allometric dataset (black), and the supplementary contribution from the 441 TLS-derived individuals (red) to the augmented allometric dataset.

Figure 7.19 presents the distribution of stems, by H, with $H \geq 20\,\text{m}$, across the Chave et al. (2014) allometric dataset. Overlaid in this figure is the contribution to the augmented allometric dataset by the TLS data. It can be seen, that in total, this supplementation has increased the number of individuals with $H \geq 20\,\text{m}$, from 1036 to

1477 This results in the composition of individuals with $H \geq 20\,\text{m}$ across the allometric dataset rising from 25.9 % to 33.2 %.

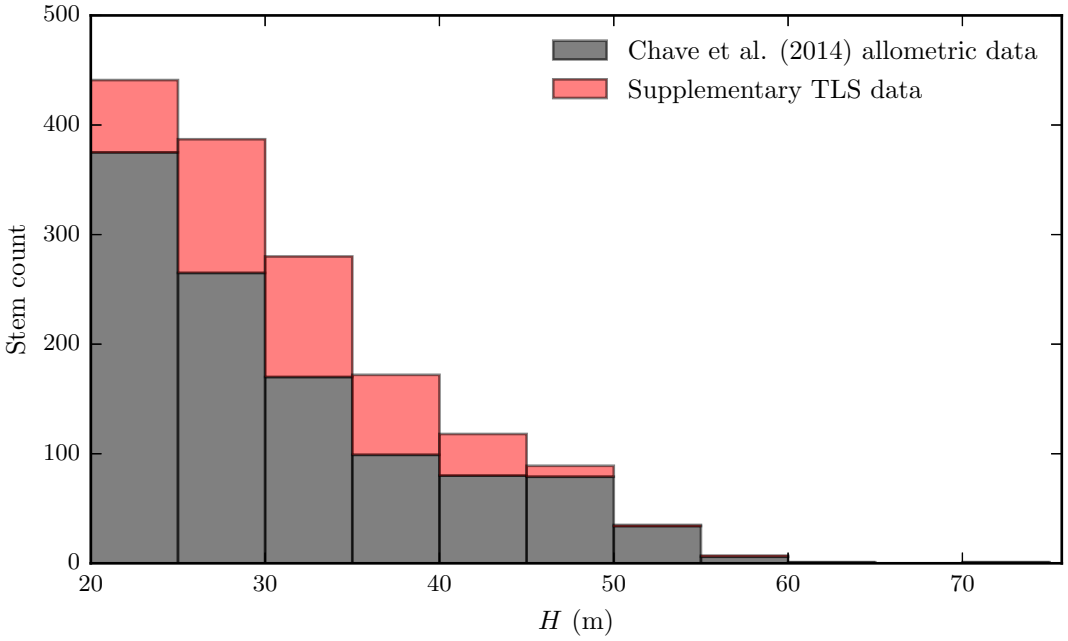


Figure 7.19: The distribution of stems, by H, with $H \geq 20\,\text{m}$ across the Chave et al. (2014) allometric dataset (black), and the supplementary contribution from the 441 TLS-derived individuals (red) to the augmented allometric dataset.

This new pan-tropical allometric model is then constructed on the augmented allometric dataset through the same methods presented in chapter 2. The model is a non-linear least squares fit through the allometric dataset, employing the single amalgamated allometric regressor, $D^2H\rho$. The tree-level uncertainty, U_t, comprises the tree-level allometric and measurement uncertainties, that are quantified through non-linear quantile regression and Monte Carlo simulation respectively. The plot-level uncertainty, P_t, comprises the plot-level allometric and measurement uncertainties, that are quantified through the wild bootstrap and Monte Carlo simulation respectively. The primary assumption that is made here is that the TLS-derived estimates of AGB_{TLS}, D and H are error-free.

This new NLS pan-tropical allometric model, derived from the augmented allometric dataset, takes the form:

$$\text{AGB} = 0.403(D^2H\rho)^{1.024} + 53.394 \tag{7.10}$$

The model, and the corresponding 95 % confidence and prediction intervals, are presented in figure 7.20. Overlaid in this figure, is the original model and confidence and prediction intervals constructed from the original Chave et al. (2014) allometric dataset.

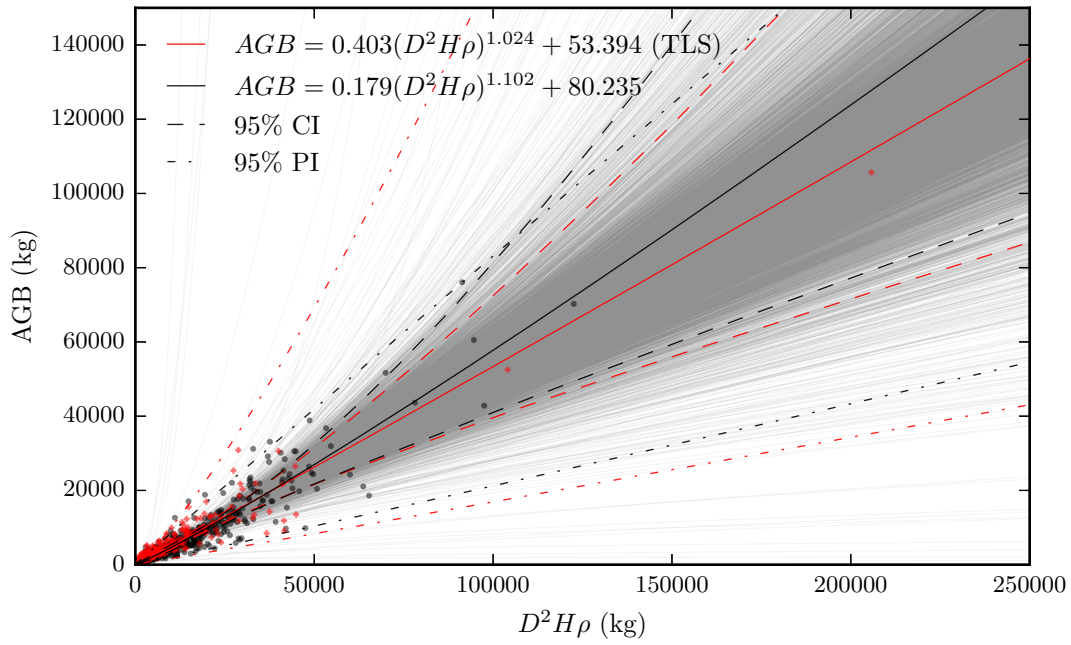


Figure 7.20: In black, the NLS model and 95 % confidence and prediction intervals constructed from the original Chave et al. (2014) allometric dataset; in red, the NLS model and 95 % confidence and prediction intervals constructed from the Chave et al. (2014) allometric dataset that has been supplemented by the 441 TLS-derived stems.

It can be seen that across values of $D^2H\rho$ below approximately 50 000 kg, the new model predicts larger estimates of AGB with respect to the original model. However, above this value, the inverse it true, such that for higher values of $D^2H\rho$, the new model predicts lower estimates of AGB with respect to the original model. It can also be seen that the 95 % confidence intervals, generated via the wild bootstrap, are quite significantly narrower than the original model. However, the change in 95 % prediction intervals is more variable. Across the value of $D^2H\rho$ below approximately 10 000 kg, the prediction intervals are narrower than their counterparts from the original model. After this value, and particularly so for the upper interval, there is significant widening when compared with the original model.

In chapter 2, the NLS model and the uncertainty quantification methods were applied to field data acquired from four field sites: ANK-01, LPG-01, MNG-03 and MNG-04. Table 7.6 presents the results from the application of the new model and compares them with the original model.

Plot code	AGB (kg)	$AGB_{CI_{lower}}$ (kg)	$AGB_{CI_{upper}}$ (kg)	U_p (%)	$\bar{U}_t$ (%)
Chave et al. (2014) allometric data					
NLS model: $AGB = 0.179(D^2H\rho)^{1.102} + 80.235$					
ANK-01	300 089	222 050	385 346	27.2	73.9
LPG-01	470 224	347 302	634 450	30.5	77.1
MNG-03	525 984	415 776	636 325	21.0	80.3
MNG-04	379 779	289 652	479 437	25.0	81.3
Chave et al. (2014) allometric data + supplementary TLS data					
NLS model: $AGB = 0.403(D^2H\rho)^{1.024} + 53.394$					
ANK-01	335 220	258 943	416 822	23.5	51.7
LPG-01	481 798	373 879	617 992	25.3	55.4
MNG-03	574 136	468 087	679 025	18.4	60.2
MNG-04	422 537	334 203	515 150	21.4	58.8

Table 7.6: Comparison between the plot-scale AGB and uncertainty, across four sets of considered field data, for the NLS model constructed from the original Chave et al. (2014) allometric dataset (upper), and the NLS model constructed from the Chave et al. (2014) allometric dataset that has been supplemented by the 441 TLS-derived stems (lower).

It can be seen that the tree-level uncertainty has reduced from an average of 78.2 % to 56.5 %. This is purely due to the narrower prediction intervals across the values of $D^2H\rho$ below 10 000 kg. The widening of the prediction intervals for higher values of $D^2H\rho$ will result in U_t across larger stems being significantly larger than that derived from the original model.

At the plot-scale, across the four plots, when comparing the old model to the new model, total AGB increased from 1 676 076 kg to 1 813 691 kg, representing a percentage difference of 7.89 %. When comparing the two models, it can be seen that U_p has marginally reduced through the introduction of this new model. Averaged across the four plots, U_p reduced from 25.9 % to 22.2 %. This is due to the reduced upper confidence interval in the new model that is seen across the full distribution of $D^2H\rho$. However, fundamentally it still remains the case that of the four plots considered only the smallest plot of 335 220 kg and largest plot of 574 136 kg, were statistically distinguishable from one another.

7.4 Discussion

7.4.1 Comparison of the TLS and allometric estimates of AGB

Chapter 2 identified through a novel experimental study that the large uncertainties in pan-tropical allometric models make them unfit for purpose, resulting in the urgent requirement for the introduction of alternative methods. The first objective of this

chapter was to demonstrate that new TLS-derived methods are capable of fulfilling this requirement. This demonstration has been through application of novel algorithm to TLS data acquired from 3.25 ha of predominately tropical forest. From these data, estimates of tree-scale AGB for 441 individuals, and plot-scale AGB for four forest plots, has been estimated. These results then, have demonstrated for the first time, the wide-scale application of TLS data to tree- and plot-scale AGB estimation, making these new TLS-derived methods a viable alternative to allometrics.

From these novel data, the next objective of this chapter was to compare the new TLS-derived estimates of AGB with their allometric counterparts. This is the first time allometric estimates of pan-tropical tree-scale AGB have been independently assessed through methods other than destructive harvest. The principal result derived from this comparison is that across the 441 considered trees, the TLS-derived estimate of AGB was 1 740 351 kg, whilst the allometric estimate was 1 454 067 kg. This represents an overall percentage difference between the two methods of 17.9 %. It should be noted that this result is broadly in line with the results from the destructively harvested data presented in chapter 6; where it was shown that across 50 *Eucalyptus* spp. individuals, that the divergence between the TLS-derived and allometric estimates was 30.2 % (with the TLS and allometric estimates a 7.9 % overestimate and 20.5 % underestimates respectively). Across the plots KARA-001, LPG-01, CAX-A and NOU-11, the percentage difference between the TLS- and allometric-derived estimates of plot-scale AGB was 34.5 %, 16.4 %, 5.3 % and 28.5 % respectively. This illustrates quite a substantial variance across the considered plots, although in all instances, the TLS-derived estimates of AGB were larger than the associated allometric estimates.

This finding is potentially very significant, as it is suggestive that current pan-tropical allometric models will underestimate tropical forest biomass. In the context of the global carbon cycle, this would suggest that the current estimate of the terrestrial vegetation carbon stock, standing at 450–650 Pg (Ciais et al. 2013), of which more than half is tropical forest (Dixon et al. 1994; IPCC 2000), is an underestimate. Consequentially, this would indicate that the current estimate of the 1.1 $\mathrm{Pg\,C\,yr^{-1}}$ sink from the atmosphere to forested land cover is poorly constrained (Pan et al. 2011). That is, the influence of deforestation, currently estimated at 7 million ha across the tropics (Williams 2003; FAO 2015; FAO 2016), is leveraged by 17.9 %. Conversely, such a result would suggest that

the potential of afforestation and reforestation to mitigate climate change by increased sequestration of atmospheric CO_2, is significantly greater than currently estimated.

7.4.2 Uncertainty in TLS-derived estimates of AGB

The principal caveat to this finding arises from the uncertainties associated with these TLS-derived estimates of tree- and plot-scale AGB. Chapter 6 presented results from TLS data acquired from 53 individuals that were subsequently destructively harvested to obtain a direct measurement of their tree-scale AGB. This permitted explicit quantification of the errors in the TLS-derived estimates of tree- and plot-scale AGB, which were shown to be 23.3 % and 7.9 % respectively. However, these validation data were acquired from *Eucalyptus* spp. open forest in Rushworth State Forest, Australia and *Quercus* spp. woodland in Alice Holt Forest, UK. As such, questions must be asked as to the transferability of these errors if they are to act as a generalisation for the uncertainties attributed to TLS-derived estimates of AGB; especially so when the TLS data are acquired from markedly different forest types and species. The sources of uncertainty introduced by these new TLS-derived methods, can be summarised as:

i) The incomplete or non-uniform sampling of the tree.

ii) Erroneous segmentation of the tree-level point cloud by the *treeseg* algorithm.

iii) Losses in the tree-level point cloud information content arising from downsampling.

iv) Misspecification of the input parameters into the Raumonen et al. (2013) shape fitting algorithm by the automatic QSM optimisation routine.

v) Erroneous cover generation, segmentation and cylinder fitting by the Raumonen et al. (2013) shape fitting algorithm.

vi) Misrepresentation of the local surface topography of the tree by a cylinder.

vii) Interference of leaf returns in the tree-level point cloud on the reconstruction of woody structure.

viii) Incorrect definition of tree-scale average wood density.

ix) Ignoring non-constant wood density through the tree.

x) Ignoring the bark fraction.

xi) Proportion of heart rot permeating the stem.

Speculatively, it is assumed here that the errors in the validation data arising from sources ii), iii), iv) and v) will be broadly transferable to other TLS data. That is, the uncertainties attributed to the modelling algorithms should be largely invariant of forest type and species. This is due to the underlying design principles of the *treeseg* algorithm, the Raumonen et al. (2013) shape-fitting algorithm and the automatic QSM optimisation routine to be as generic as possible, so as to not be species-specific constrained. The remaining seven sources of uncertainty stem from the underlying structural and taxonomic composition of the considered forest plots.

Uncertainty source i), the incomplete or non-uniform TLS sampling of plot is almost entirely due to data acquisition protocol and the structural complexity of the forest plot. Interestingly, the 95 % confidence intervals around each estimate of TLS-derived tree-scale AGB varied across the plots. These were derived from the standard deviation of tree-level volume across the 20 QSMs in the optimised parameter set. Across the three tropical forest plots LPG-01, CAX-A and NOU-11, the mean size of these intervals were 20.0 %, 18.9 % and 13.8 % respectively. The observed reduction correlates directly with successive improvements in the sampling regime. For instance, in plot LPG-01, the sampling regime employed a 20 m systematic square grid using a 0.06° angular sampling resolution. In plot CAX-A, a 20 m systematic square grid was used with a 0.04° angular sampling resolution. In plot NOU-11, a 10 m systematic square grid was used with a 0.04° angular sampling resolution. That is, with each subsequent improvement in the sampling regime, there was an associated reduction in the modelling uncertainty. This was most pronounced in the transition from a 20 m systematic square grid to a 10 m systematic square grid, where the number of scan locations increased from 36 to 121. Inevitably however, because all TLS viewpoints are from close to the forest floor, the influence of increasing occlusion, step-distance between consecutively fired pulses and beam footprint, universally result in incomplete and non-uniform sampling. Reducing the influence of this uncertainty on TLS-derived estimates of tree-scale AGB will primarily come from the introduction of new TLS instruments with smaller beam divergences.

Uncertainty source vi), the misrepresentation of the local surface topography of the tree by a cylinder, is potentially one of the largest sources of error in these TLS-derived

estimates of tree-scale AGB. This is primarily due to the presence, across the tropical forest plots, of stems with buttress roots. Of particular note were figures 7.7c, 7.11d, 7.15a and 7.15d that presented four QSMs overlaid on the tree-level point clouds from which they were constructed; in each of the four instances, buttress formations were present. Across these examples, the cylinder in the QSM representing the buttress section of the stem was a poor representation of the local surface topography. However, in all but figure 7.7c, where a complex above-surface rooting system was present, the cylinder model appeared underdeveloped. That is, the expectation for these three examples would be for tree-scale AGB to increase, that would, in turn, increase the divergence between the TLS and allometric estimates of AGB. One solution that is currently being researched to reduce this error is to reconstruct surface topography via a mesh (Smith et al. 2014), as opposed to a geometric primitive. Unlike cylinders, a mesh can represent any shape, although such methods are usually reliant upon a consistent sampling of the tree surface, something that is not always feasible in point clouds from tropical forests.

Uncertainty source vii), the interference of leaf returns in the tree-level point cloud on the reconstruction of woody structure is the sole uncertainty that can be explicitly assessed by the validation data considered in chapter 6. Pre-harvest, the 3 *Quercus* spp. individuals from Alice Holt Forest, UK, were scanned in leaf-off and leaf-on conditions, where it was shown that the explicit mean error in the TLS-derived estimate of tree-scale AGB increased from 3.7 % to 26.2 % respectively. This put the tree-scale error largely in line with that observed across the validation data from Rushworth State Forest, Australia; demonstrating that the equal treatment of leaf returns in the tree-level point cloud does significantly impact on the retrieval of tree-scale AGB. It should be noted however, that across the validation data from Rushworth State Forest, Australia, that at the plot-scale, this error reduced down to 7.9 %. It should also be noted that across the 3 *Quercus* spp. individuals, branching occurred at a relatively low height along the stem, resulting in significant quantities of leafy material across the majority of the extent of each tree. For the tropical trees considered here, whose architecture is typically driven through light competition (Coomes et al. 2011), where the leafy material predominately resides in the crown, it is anticipated that the influence of leaf returns on the reconstruction would be reduced. Looking forward, there are two avenues for reducing the impact of leaf interference on the reconstruction of woody structure. The first avenue is

to further develop algorithms to remove leaf returns, such as the method developed in chapter 5 section 5.3.9, that was shown to permit a reasonable classification of wood and leaf returns. The other avenue is the introduction of TLS instruments that are capable of discriminating wood and leaf returns, such as the dual wavelength instruments reviewed in chapter 3 section 3.1.6.

Uncertainty source viii), the incorrect definition of tree-scale average wood density, will exert two influences over the error in TLS-derived tree-scale AGB. In the absolute sense, the use of the plot-level average wood density induces an error in the tree-scale estimate of AGB that will propagate through to the plot-scale estimate. For instance, across plot LPG-01, the tree-level average wood density ranged from 378–918 $\mathrm{kg\,m^{-3}}$. However, as the plot-level average wood density considered across all the plots is weighted by basal area, the impact of this should be reasonably low. In a relative sense, the use of the plot-level average wood density will influence the percentage difference between the TLS and allometric estimates of AGB. This is because the relationship with wood density, between the two methods, varies, in line with:

$$\mathrm{AGB_{TLS}} \propto \rho \tag{7.11}$$

$$\mathrm{AGB_{Allom}} \propto \rho^{1.102} \tag{7.12}$$

However, as the allometric exponent is reasonably close to 1, the influence of this should be minimal. To reduce the uncertainty arising from the incorrect definition of tree-scale average wood density, reliable methods need to be introduced to match the TLS data to the field data. This would permit a value of tree-level wood density to be attributed to each tree-level point cloud.

Uncertainty source ix), is the non-consideration of variation in wood density through each tree. That is, a constant tree-level wood density has been applied in the TLS-derived estimates of tree-scale AGB. It has been demonstrated in the literature that tree-level wood density varies through a tree, both radially and with height (Knapic et al. 2008). It has also been shown that the magnitude of this response is species-specific (Nock et al. 2009). As such, with so little data describing this variation available, explicit quantification of this uncertainty will be particularly difficult.

Uncertainty source x), is the disregard of the bark fraction. The contents of each

cylinder inside the QSM are considered wood, such that a constant tree-level wood density has been applied. As a result, there is no accommodation for fraction of bark contributing to the cylinder, which will invariably have a different density to the wood. The bark fraction of a tree is a complex relationship between species and size class. It is also thought to be related to bio-mechanical support, respiration and defence from the environment (Paine et al. 2010). With the addition of bark fraction varying by height, the quantification of this uncertainty will also be difficult.

The final uncertainty source xi), is the proportion of heart rot permeating the stem. The contents of each cylinder inside the QSM are considered to be healthy, living wood. It is known, although the body of literature is very small, that the insides of trees can be hollow, due to rot (Lee et al. 1999). Therefore, the presence of heart rot in a tree would lead to an overestimate in the TLS-derived estimate of tree-scale AGB. The magnitude of the hollowing can vary significantly on a tree-by-tree basis, that currently, is impossible to assess. One interesting new development in this area is the used of sonic tomography, a non-destructive remote sensing technique that permits an evaluation of internal wood health (Gilbert et al. 2016).

This section has explicitly defined the sources of uncertainty attributed to these new TLS-derived estimates of tree-scale AGB, although no attempt has been made to explicitly quantify them. The application of the TLS methods to the validation data described in chapter 6, permitted an estimation of error that implicitly incorporated a number of these sources. As such, these value of 23.3 % and 7.9 % at the tree- and plot-scale respectively, present the current best estimate of the uncertainties attributed to these new TLS-derived estimates of AGB. However, care must be taken when using these errors as a generalisation of the uncertainties that can be expected to accompany TLS-derived estimates of AGB. This is particularly true across TLS data acquired from tropical forests that tend to be so species-rich and structurally variable. Whilst further work has been suggested to account for some of these uncertainties, ultimately, further destructive harvest data is required as this represents the gold standard for method validation. This should be made a priority in light of the large discrepancies observed between the TLS- and allometric-derived estimates of tree- and plot-scale AGB.

7.4.3 Uncertainty in allometric estimates of AGB

The key result from this chapter was the observed 17.9 % difference between the TLS- and allometric-derived estimates of tree-scale AGB across the considered 441 individuals. One final, fundamental caveat to the interpretation of this difference, is the consideration of the uncertainties associated with the allometrics. It was shown in table 7.6, that the uncertainties in allometric estimates of plot-scale AGB averaged 25.9 %. If, for a moment, the TLS-derived estimates of AGB are considered error-free, then because of these large allometric uncertainties, the observed percentage difference between the two methods cannot be considered statistically significant. As such, any hypotheses as to why the allometric estimates tend to produce estimates of tree-scale AGB below their TLS-derived counterparts, are statistically meaningless.

Interestingly, the new pan-tropical allometric model that was constructed on the augmented Chave et al. (2014) allometric dataset, reduced the tree-scale uncertainties from 78.2 % to 56.5 % and the plot-scale uncertainties from 25.9 % to 22.2 %. This might suggest that there is scope for additional destructive harvest data to further constrain the allometric models. However the core argument against this persists, that across the four considered field sites, it remained the case that only the plot with the smallest plot-scale AGB (335 220 kg) and largest plot-scale AGB (574 136 kg) were statistically distinguishable from one another. As such, the conclusion of chapter 2 still remains valid that the large uncertainties associated with allometric models make them not fit for purpose. This is fundamentally due to the inability of the allometric regressors, D, H and ρ, to capture the observed pan-tropical variance in AGB.

7.5 Conclusions

This chapter has demonstrated, for the first time, that the new methods developed and applied in the work for this thesis, permit TLS-derived estimates of tree- and plot-scale AGB to be considered a viable alternative to allometric methods. This demonstration has been achieved by applying these new methods to TLS data acquired from the 3.25 ha of predominately tropical forest described in chapter 4. From these data, estimates of tree-scale AGB for 441 individuals, and plot-scale AGB for four forest plots, has been retrieved.

Following this retrieval, these new TLS-derived estimates have been compared with

their allometric counterparts, where it has been shown that the total AGB was estimated at 1 740 351 kg and 1 454 067 kg respectively. This represents an overall percentage difference between the two methods of 17.9 %. These results have potentially large implications on the global carbon cycle, as it indicates that current pan-tropical allometric models will significantly underestimate tropical forest carbon storage. Consequentially, the carbon sink from the atmosphere to forested land cover, a key mitigant of climate change, is more potently affected by deforestation and afforestation/reforestation than currently estimated.

Finally, this chapter concluded by presenting a new NLS pan-tropical allometric model that was constructed on the Chave et al. (2014) allometric dataset supplemented by these new TLS data. Despite an observed reduction in the uncertainties when compared to the original NLS model, it was demonstrated that large swathes of tropical forest remain statistically indistinguishable from one another. As a result, the conclusion of chapter 2, that the NLS pan-tropical allometric model was not fit for purpose, remains valid.

8 Conclusions

8.1 Thesis overview

This thesis has presented two new methods for estimating tree- and plot-scale above-ground biomass (AGB). The first approach, presented in chapter 2, built on the current allometric methods relating simple measurements of tree structure, to AGB. An experimental study was presented that critiqued the foremost pan-tropical allometric models, with particular focus on both the statistical methods used in their construction, and on the characterisation of uncertainty in their estimates of tree- and plot-scale AGB.

This study also proposed a new non-linear pan-tropical allometric model. Beside this model, a novel framework was developed for calculating uncertainty in allometric-derived estimates of tree- and plot-scale AGB through non-parametric methods. The study was concluded by applying these new methods to field data from four censused 1 ha tropical forest plots.

The second approach was the estimation of tree- and plot-scale AGB from new 3D measurements of forest structure captured from terrestrial laser scanning (TLS). This thesis had the primary objective of demonstrating that tree- and plot-scale estimates of AGB could be accurately retrieved on a wide-scale from TLS point clouds. Chapter 3 identified three knowledge gaps that currently prohibit the wide-scale application of TLS data for this purpose. These knowledge gaps were bridged in this thesis through the development and application of a stack of novel software.

Chapter 5 presented the *treeseg* algorithm, developed here to automatically extract tree-level point clouds from plot-level point clouds. Chapter 6 then introduced quantitative structural models (QSM), 3D descriptions of woody tree structure reconstructed from tree-level point clouds, through which tree-scale AGB can be retrieved via volume estimation. This chapter presented new methods to automate the construction of QSMs generated from the Raumonen et al. (2013) shape fitting algorithm.

These new methods were then validated in chapter 6 by applying them to TLS data acquired from 53 individuals that were subsequently destructively harvested to obtain a direct measurement of tree-scale AGB. Finally, these methods were then applied in

chapter 7 to data acquired from 3.25 ha of predominately tropical forest, the first data of this kind, to retrieve estimates of tree and plot-scale AGB.

8.2 Key findings

The allometric modelling critique presented in chapter 2 demonstrated that the sole statistical method used in the construction of all widely used pan-tropical allometric models, namely log-transformed ordinary least squares linear regression, is inappropriate. This was due to the models not conforming to the underlying assumptions of homoscedasticity and normality. A consequence of this is the return of ill-conditioned error terms that result in spurious statistical inferences being drawn from the models. Furthermore, a reliance by these models on the error terms for re-transformation into arithmetic-space, renders predictive inferences biased.

In contrast, the non-linear model proposed in this thesis permitted unbiased estimates of AGB because it was constructed in arithmetic-space, whilst valid statistical inference was derived through non-parametric methods. When applied to the four considered tropical forest plots, it was demonstrated that relative uncertainty in tree- and plot-scale estimates of AGB exceeded 75 % and 25 % respectively. Of the four plots considered, only the smallest and largest, with an estimated AGB of 300 089 kg and 525 983 kg respectively, were statistically distinguishable from one another.

The implication of such large uncertainties is that the AGB of large swathes of tropical forest will be statistically indistinguishable when inferred through this non-linear pan-tropical allometric model. A further implication is that temporal change detection is permissible in only the most extreme of disturbances. Due to these large uncertainties, it was therefore concluded that the non-linear model was not fit for purpose.

This conclusion suggests that there is a need to re-evaluate the validity and uncertainty of AGB products at all scales because of the current ubiquitous dependency of up-scaled products on allometry. This conclusion also highlighted the requirement for the introduction of alternative methods that are capable of more accurately estimating tree- and plot-scale AGB.

It has been demonstrated in this thesis that TLS-derived methods provide a viable alternative to traditional allometry. The amalgamation of the *treeseg* algorithm, the Raumonen et al. (2013) shape fitting algorithm and the QSM optimisation routine,

has provided, for the first time, a near-automatic approach for the wide-scale retrieval of tree and plot-scale AGB from TLS data. When these methods were applied to the TLS data acquired from 53 destructively harvested individuals, it was demonstrated that relative error in the estimates of tree- and plot-scale AGB was 23.3 % and 7.9 % respectively. These errors in the TLS-derived methods are substantially lower than the expected uncertainty in their allometric counterparts.

When these methods were applied to the TLS data collected from 3.25 ha of forest, 441 tree level-point clouds were extracted. Of these stems, 324 were extracted entirely automatically, representing a success rate of 77.6 %; while the remaining 117 stems required only moderate manual refinement. The vast majority of these data were collected in structurally complex tropical forest, so the expectation is for these new methods to be readily transferable to different forest types.

These new TLS-derived estimates of tree-scale AGB were then directly compared against allometric-derived estimates. It was shown that the TLS- and allometric-derived estimates of total AGB were 1 740 351 kg and 1 454 067 kg respectively. This represented an overall percentage difference between the two methods of 17.9 %.

These results suggest that pan-tropical allometric models may significantly underestimate tropical forest carbon storage. The implication of this result is that the current expectation of the impact of deforestation and afforestation/reforestation on the terrestrial carbon sink, a key mitigant of climate change, is leveraged. This reinforces the conclusion that there is a need to re-evaluate the validity and uncertainty of pan-tropical AGB products at all scales, as their accuracy is vital for informed pan-tropical forest management and conservation activities.

8.3 Future work

8.3.1 Allometric methods

A number of sources of uncertainty were omitted during the development of the non-parametric uncertainty quantification framework. Most notable was the ecological variance uncertainty, that expressed population-scale variability not captured by the allometric dataset (i.e., the variation of the AGB of each tree across the tropics with respect to the allometric regressors). Whilst the most comprehensive allometric dataset available was used here, it is highly unlikely to be of sufficient size, species or geographical range to

capture this population-scale variability. The impact of this uncertainty on allometric-derived estimates of pan-tropical tree- and plot-scale AGB is unknown, but it would be reasonable to speculate that it would have some material impact. One interesting avenue for further work would be to jackknife the allometric dataset by spatial, species and size-class distribution, so as to explore the influence of these individual factors on total uncertainty.

The measurement uncertainty (i.e., uncertainty introduced during the collection of the allometric regressors) was also largely discarded because it was assumed that the errors would conform to a normal distribution. Due to this assumption, across a 1 ha plot containing several hundred stems, the measurement uncertainty will largely average out. Speculatively, it would be highly unlikely that these measurement errors would conform to such assumptions. For example, when measuring individuals in the emergent layer, it would be near-impossible to overestimate tree height, which excludes the possibility of this measurement error being normally distributed. The Monte Carlo simulation approach presented in this thesis to characterise measurement uncertainty would readily permit further work into investigating the response of uncertainty in allometric-derived estimates of tree- and plot-scale AGB, to alternative measurement error distributions.

8.3.2 TLS methods

The development and application of new algorithms in this thesis presented a method to near-automatically retrieve tree- and plot-scale AGB from TLS data. The first (of two) limitations of these methods were that only stems with $D \geq 0.2\,\text{m}$ were considered for extraction from the plot-level point clouds. In this thesis, this was justified because the interest lay with tree- and plot-scale AGB retrieval, where stems with $D < 0.2\,\text{m}$ generally only contribute 5–10 % to plot-scale AGB. However, for more broad applications, and in line with traditional field measurement protocol, all stems with $D \geq 0.1\,\text{m}$ need to be extracted from the plot-level point clouds. There is no obvious limitation as to why the *treeseg* algorithm could not be successfully applied to smaller stems, it was just not considered in the work for this thesis.

The second limitation is that some tree-level point clouds extracted from the *treeseg* algorithm required manual refinement. Further work should be undertaken to increase the extraction success rate through optimisation of the cluster growing algorithm; one

interesting approach to this would be the introduction of supervised machine learning.

Whilst validation of these TLS-derived methods and results has been been undertaken in this thesis, the validation data themselves are somewhat limited. That is, these data were neither acquired from tropical forests, or from particularly large stems. Further work should prioritise the acquisition of destructive harvest data to permit a more comprehensive understanding of how these error in the TLS-derived methods permeate across varied forest types and structures. Further to this, it would interesting to use such data to explicitly explore the individual sources of error that were described in chapter 7, such as the impact of the inclusion of leaf returns on woody reconstruction, or the disregard of bark fraction, variable wood density and trunk hollowing.

The comparison between the TLS- and allometric- derived estimates of AGB presented a result that has a potentially significant implication for the global carbon cycle. Although the TLS data presented here were acquired from tropical forest from two continents, a sample of three 1 ha plots from nominally the same forest type (moist, Terra Firme, lowland, mixed species, old-growth) is a limitation. Further work should concentrate on the acquisition of TLS data from a broad range of locations and forest types to determine the global impact of discrepancies between these TLS- and allometric-derived estimates via up-scaling.

Finally, the application of the new 3D measurements of forest structure presented in this thesis has been to estimation of AGB. The methods presented in this thesis provide access to unparalleled wide-scale descriptions of 3D tree structure. Further work could see these descriptions applied broadly across the discipline of environmental sciences; ranging from application in the testing of fundamental ecological hypotheses, to the calibration and validation of forthcoming spacebourne missions.

References

Ahmed, R., Siqueira, P., Hensley, S., and Bergen, K. (2013). "Uncertainty of forest biomass estimates in north temperate forests due to allometry: implications for remote sensing". In: *Remote Sensing* 5 (6), pp. 3007–3036. DOI: 10.3390/rs5063007.

Åkerblom, M., Raumonen, P., Kaasalainen, M., and Casella, E. (2015). "Analysis of geometric primitives in Quantitative Structure Models of tree stems". In: *Remote Sensing* 7 (4), pp. 4581–4603. DOI: 10.3390/rs70404581.

Anderson, T. W. and Darling, D. A. (1954). "A test of goodness of fit". In: *Journal of the American Statistical Association* 49 (268), pp. 765–769. DOI: 10.1080/01621459.1954.10501232.

Avitabile, V., Herold, M., Heuvelink, G. B. M., Lewis, S. L., Phillips, O. L., et al. (2016). "An integrated pan-tropical biomass map using multiple reference datasets". In: *Global Change Biology* 22 (4), pp. 1406–1420. DOI: 10.1111/gcb.13139.

Baccini, A., Goetz, S. J., Walker, W. S., Laporte, N. T., Sun, M., et al. (2012). "Estimated carbon dioxide emissions from tropical deforestation improved by carbon-density maps". In: *Nature Climate Change* 2 (3), pp. 182–185. DOI: 10.1038/nclimate1354.

Baker, T. R., Phillips, O. L., Malhi, Y., Almeida, S., Arroyo, L., et al. (2004). "Variation in wood density determines spatial patterns in Amazonian forest biomass". In: *Global Change Biology* 10 (5), pp. 545–562. DOI: 10.1111/j.1365-2486.2004.00751.x.

Baskerville, G. L. (1972). "Use of logarithmic regression in the estimation of plant biomass". In: *Canadian Journal of Forest Research* 2 (1), pp. 49–53. DOI: 10.1139/x72-009.

Bauwens, S., Bartholomeus, H., Calders, K., and Lejeune, P. (2016). "Forest inventory with terrestrial LiDAR: a comparison of static and hand-held mobile laser scanning". In: *Forests* 7 (6). 127. DOI: 10.3390/f7060127.

Bawa, K. S. (1990). "Plant-Pollinator Interactions in Tropical Rain Forests". In: *Annual Review of Ecology and Systematics* 21, pp. 399–422. DOI: 10.1146/annurev.es.21.110190.002151.

Belsley, D. A., Kuh, E., and Welsch, R. E. (1980). *Regression diagnostics: identifying influential data and source of collinearity*. John Wiley & Sons, Inc. ISBN: 978-0-471-05856-4. DOI: 10.1002/0471725153.

Bonan, G. B. (2008). "Forests and climate change: forcings, feedbacks, and the climate benefits of forests". In: *Science* 320 (5882), pp. 1444–1449. DOI: 10.1126/science.1155121.

Bréda, N. J. J. (2003). "Ground-based measurements of leaf area index: a review of methods, instruments and current controversies". In: *Journal of Experimental Botany* 54 (392), pp. 2403–2417. DOI: 10.1093/jxb/erg263.

Breusch, T. S. and Pagan, A. R. (1979). "A simple test for heteroscedasticity and random coefficient variation". In: *Econometrica* 47 (5), pp. 1287–1294. DOI: 10.2307/1911963.

Brown, J. H., Gillooly, J. F., Allen, A. P., Savage, V. M., and West, G. B. (2004). "Toward a metabolic theory of ecology". In: *Ecology* 85 (7), pp. 1771–1789. DOI: 10.1890/03-9000.

Brown, S. (1997). *Estimating biomass and biomass change of tropical forests: a primer*. Food and Agriculture Organization of the United Nations. ISBN: 978-92-5-103955-7.

Bruijnzeel, L. A. (2004). "Hydrological functions of tropical forests: not seeing the soil for the trees?" In: *Agriculture, Ecosystems and Environment* 104 (1), pp. 185–228. DOI: 10.1016/j.agee.2004.01.015.

Bryant, D., Nielsen, D., and Tangley, L. (1997). *The last frontier forests: ecosystems and economies on the edge*. World Resources Institute. ISBN: 978-1-56973-198-7.

Calders, K., Armston, J., Newnham, G., Herold, M., and Goodwin, N. (2014). "Implications of sensor configuration and topography on vertical plant profiles derived from terrestrial LiDAR". In: *Agricultural and Forest Meteorology* 194, pp. 104–117. DOI: 10.1016/j.agrformet.2014.03.022.

Calders, K., Newnham, G., Burt, A., Murphy, S., Raumonen, P., et al. (2015). "Nondestructive estimates of above-ground biomass using terrestrial laser scanning". In: *Methods in Ecology and Evolution* 6 (2), pp. 198–208. DOI: 10.1111/2041-210X.12301.

Canadell, J. G. and Raupach, M. R. (2008). "Managing forests for climate change mitigation". In: *Science* 320 (5882), pp. 1456–1457. DOI: 10.1126/science.1155458.

Carlson, T. N. and Ripley, D. A. (1997). "On the relation between NDVI, fractional vegetation cover, and leaf area index". In: *Remote Sensing of Environment* 62 (3), pp. 241–252. DOI: 10.1016/S0034-4257(97)00104-1.

Chave, J., Andalo, C., Brown, S., Cairns, M. A., Chambers, J. Q., et al. (2005). "Tree allometry and improved estimation of carbon stocks and balance in tropical forests". In: *Oecologia* 145 (1), pp. 87–99. DOI: 10.1007/s00442-005-0100-x.

Chave, J., Condit, R., Aguilar, S., Hernandez, A., Lao, S., et al. (2004). "Error propagation and scaling for tropical forest biomass estimates". In: *Philosophical Transactions of the Royal Society of London B: Biological Sciences* 359 (1443), pp. 409–420. DOI: 10.1098/rstb.2003.1425.

Chave, J., Réjou-Méchain, M., Búrquez, A., Chidumayo, E., Colgan, M. S., et al. (2014). "Improved allometric models to estimate the aboveground biomass of tropical trees". In: *Global Change Biology* 20 (10), pp. 3177–3190. DOI: 10.1111/gcb.12629.

Ciais, P., Sabine, C., Bala, G., Bopp, L., Brovkin, V., et al. (2013). "Carbon and Other Biogeochemical Cycles. Contribution of working group I to the fifth assessment report of the Intergovernmental Panel on Climate Change". In: *Climate change 2013: the physical science basis*. Cambridge University Press. ISBN: 978-1-107-66182-0. DOI: 10.1017/CBO9781107415324.015.

Clark, D. A., Brown, S., Kicklighter, D. W., and Chambers, J. Q. (2001). "Measuring net primary production in forests: concepts and field methods". In: *Ecological Applications* 11 (2), pp. 356–70. DOI: 10.1890/1051-0761(2001)011[0356:MNPPIF]2.0.CO;2.

Clark, D. B. and Clark, D. A. (2000). "Landscape-scale variation in forest structure and biomass in a tropical rain forest". In: *Forest Ecology and Management* 137 (1), pp. 185–198. DOI: 10.1016/S0378-1127(99)00327-8.

Clark, D. B. and Kellner, J. R. (2012). "Tropical forest biomass estimation and the fallacy of misplaced concreteness". In: *Journal of Vegetation Science* 23 (6), pp. 1191–1196. DOI: 10.1111/j.1654-1103.2012.01471.x.

Clark, R., Swayze, G., Wise, R. A., Livo, K. E., Hoefen, T. M., et al. (2007). *USGS digital spectral library splib06a: U.S. Geological Survey, digital data series 231*. URL: http://speclab.cr.usgs.gov/spectral.lib06 (visited on 2017-07-21).

Coomes, D. A. and Allen, R. B. (2009). "Testing the Metabolic Scaling Theory of tree growth". In: *Journal of Ecology* 97 (6), pp. 1369–1373. DOI: 10.1111/j.1365-2745.2009.01571.x.

Coomes, D. A., Lines, E., and Allen, R. B. (2011). "Moving on from Metabolic Scaling Theory: hierarchical models of tree growth and asymmetric competition for light". In: *Journal of Ecology* 99 (3), pp. 748–756. DOI: 10.1111/j.1365-2745.2011.01811.x.

Costanza, R., Groot, R. de, Sutton, P., Ploeg, S. van der, Anderson, S. J., et al. (2014). "Changes in the global value of ecosystem services". In: *Global Environmental Change* 26, pp. 152–158. DOI: 10.1016/j.gloenvcha.2014.04.002.

Côté, J.-F., Widlowski, J.-L., Fournier, R. A., and Verstraete, M. M. (2009). "The structural and radiative consistency of three-dimensional tree reconstructions from terrestrial LiDAR". In: *Remote Sensing of Environment* 113 (5), pp. 1067–1081. DOI: 10.1016/j.rse.2009.01.017.

Crowther, T. W., Glick, H. B., Covey, K. R., Bettigole, C., Maynard, D. S., et al. (2015). "Mapping tree density at a global scale". In: *Nature* 525, pp. 201–205. DOI: 10.1038/nature14967.

Cuni-Sanchez, A., White, L. J. T., Calders, K., Jeffery, K. J., Abernethy, K., et al. (2016). "African savanna-forest boundary dynamics: a 20-year study". In: *PLOS ONE* 11 (6). DOI: 10.1371/journal.pone.0156934.

D'Agostino, R. B. (1971). "An omnibus test of normality for moderate and large size samples". In: *Biometrika* 58 (2), pp. 341–348. DOI: 10.1093/biomet/58.2.341.

Danson, F. M., Gaulton, R., Armitage, R. P., Disney, M., Gunawan, O., et al. (2014). "Developing a dual-wavelength full-waveform terrestrial laser scanner to characterize forest canopy structure". In: *Agricultural and Forest Meteorology* 198, pp. 7–14. DOI: 10.1016/j.agrformet.2014.07.007.

Dassot, M., Colin, A., Santenoise, P., Fournier, M., and Constant, T. (2012). "Terrestrial laser scanning for measuring the solid wood volume, including branches, of adult standing trees in the forest environment". In: *Computers and Electronics in Agriculture* 89, pp. 86–93. DOI: 10.1016/j.compag.2012.08.005.

Disney, M., Lewis, P., and Raumonen, P. (2012). "Testing a new vegetation structure retrieval algorithm from terrestrial LiDAR scanner data using 3D models". In: *SilviLaser 2012, Vancouver, Canada, September 16–19*. URL: http://silvilaser2012.

com/wp-content/uploads/2011/11/Silvilaser2012_Full_Proceedings.pdf (visited on 2017-07-21).

Dixon, R. K., Solomon, A. M., Brown, S., Houghton, R. A., Trexier, M. C., et al. (1994). "Carbon pools and flux of global forest ecosystems". In: *Science* 263 (5144), pp. 185–190. DOI: 10.1126/science.263.5144.185.

Douglas, E. S., Martel, J., Li, Z., G., H., Hewawasam, K., et al. (2015). "Finding leaves in the forest: the dual-wavelength Echidna LiDAR". In: *IEEE Geoscience and Remote Sensing Letters* 12 (4), pp. 776–780. DOI: 10.1109/LGRS.2014.2361812.

Edwards, P. J. and Grubb, P. J. (1977). "Studies of mineral cycling in a montane rain forest in New Guinea: I. The distribution of organic matter in the vegetation and soil". In: *Journal of Ecology* 65 (3), pp. 943–969. DOI: 10.2307/2259387.

Fan, L., Smethurst, J. A., Atkinson, P. M., and Powrie, W. (2015). "Error in target-based georeferencing and registration in terrestrial laser scanning". In: *Computers and Geosciences* 83, pp. 54–64. DOI: 10.1016/j.cageo.2015.06.021.

FAO (2010). *Global forest resources assessment 2010*. Food and Agriculture Organization of the United Nations. ISBN: 978-92-5-106654-6.

FAO (2012). *State of the world's forests 2012*. Food and Agriculture Organization of the United Nations. ISBN: 978-92-5-107292-9.

FAO (2015). *Global forest resources assessment 2015*. Food and Agriculture Organization of the United Nations. ISBN: 978-92-5-109283-5.

FAO (2016). *State of the world's forests 2016*. Food and Agriculture Organization of the United Nations. ISBN: 978-92-5-109208-8.

Feldpausch, T. R., Lloyd, J., Lewis, S. L., Brienen, R. J. W., Gloor, M., et al. (2012). "Tree height integrated into pantropical forest biomass estimates". In: *Biogeosciences* 9 (8), pp. 3381–3403. DOI: 10.5194/bg-9-3381-2012.

Field, C. B., Behrenfeld, M. J., Randerson, J. T., and Falkowski, P. (1998). "Primary production of the biosphere: integrating terrestrial and oceanic components". In: *Science* 281 (5374), pp. 237–240. DOI: 10.1126/science.281.5374.237.

Fine, P. V. A., Ree, R. H., and Burnham, R. J. (2008). "The Disparity in Tree Species Richness among Tropical, Temperate, and Boreal Biomes: The Geographic Area and Age Hypothesis". In: *Tropical Forest Community Ecology*. Wiley-Blackwell. ISBN: 978-1-4051-1897-2.

Foundation, B. (2017). *Blender*. Version 2.78. URL: https://blender.org/download/ (visited on 2017-07-21).

GCOS (2010). *Implementation plan for the global observing system for climate in support of the UNFCCC (2010 update). GCOS-138 (GOOS-184, GTOS-76, WMO-TD/No. 1523)*. World Meteorological Organization. URL: http://wmo.int/pages/prog/gcos/Publications/gcos-138.pdf (visited on 2017-07-21).

Gibbs, H. K., Brown, S., Niles, J.-O., and Foley, J. A. (2007). "Monitoring and estimating tropical forest carbon stocks: making REDD a reality". In: *Environmental Research Letters* 2 (4). DOI: 10.1088/1748-9326/2/4/045023.

Gilbert, G. S., Ballesteros, J. O., Barrios-Rodriguez, C. A., Bonadies, E. F., Cedeño-Sánchez, M. L., et al. (2016). "Use of sonic tomography to detect and quantify wood decay in living trees". In: *Applications in Plant Sciences* 4 (12). DOI: 10.3732/apps.1600060.

Goetz, S. J., Baccini, A., Laporte, N. T., Johns, T., Walker, W., et al. (2009). "Mapping and monitoring carbon stocks with satellite observations: a comparison of methods". In: *Carbon Balance and Management* 4 (1). DOI: 10.1186/1750-0680-4-2.

Goldberg, D. E. and Holland, J. H. (1988). "Genetic algorithms and machine learning". In: *Machine Learning* 3 (2), pp. 95–99. DOI: 10.1023/A:1022602019183.

Gower, S. T., Kucharik, C. J., and Norman, J. M. (1999). "Direct and indirect estimation of leaf area index, fAPAR, and net primary production of terrestrial ecosystems". In: *Remote Sensing of Environment* 70 (1), pp. 29–51. DOI: 10.1016/S0034-4257(99)00056-5.

Graybill, F. A. and K., I. H. (1994). *Regression analysis: concepts and applications*. Duxbury Press. ISBN: 978-0-534-19869-5.

Greene, W. H. (2011). *Econometric Analysis*. Pearson Education. ISBN: 978-0-273-75356-8.

Hackenberg, J., Spiecker, H., Calders, K., Disney, M., and Raumonen, P. (2015). "SimpleTree - an efficient open source tool to build tree models from TLS clouds". In: *Forests* 6 (11). DOI: 10.3390/f6114245.

Hale, A. (2011). *Sapling Tree*. Version 0.2. URL: https://wiki.blender.org/index.php/Extensions:2.6/Py/Scripts/Curve/Sapling_Tree (visited on 2017-07-21).

Hallé, F., Oldeman, R. A. A., and Tomlinson, P. B. (1978). *Tropical trees and forests: an architectural analysis*. Springer-Verlag Berlin Heidelberg. ISBN: 978-3-642-81192-0. DOI: 10.1007/978-3-642-81190-6.

Hancock, S., Armston, J., Li, Z., Gaulton, R., Lewis, P., et al. (2015). "Waveform LiDAR over vegetation: an evaluation of inversion methods for estimating return energy". In: *Remote Sensing of Environment* 164, pp. 208–224. DOI: 10.1016/j.rse.2015.04.013.

Hayashi, F. (2000). *Econometrics*. Princeton University Press. ISBN: 978-0-691-01018-2.

Henning, J. G. and Radtke, P. J. (2006). "Ground-based laser imaging for assessing three-dimensional forest canopy structure". In: *Photogrammetric Engineering and Remote Sensing* 72 (12), pp. 1349–1358. DOI: 10.14358/PERS.72.12.1349.

Hopkinson, C., Chasmer, L., Young-Pow, C., and Treitz, P. (2004). "Assessing forest metrics with a ground-based scanning LiDAR". In: *Canadian Journal of Forest Research* 34 (3), pp. 573–583. DOI: 10.1139/x03-225.

Houghton, R. A. (2005). "Aboveground forest biomass and the global carbon balance". In: *Global Change Biology* 11 (6), pp. 945–958. DOI: 10.1111/j.1365-2486.2005.00955.x.

Houghton, R. A., Hall, F., and Goetz, S. J. (2009). "Importance of biomass in the global carbon cycle". In: *Journal of Geophysical Research: Biogeosciences* 114 (G2). DOI: 10.1029/2009JG000935.

Hozumi, K., Yoda, K., Kokawa, S., and Kira, T. (1969). "Production ecology of tropical rain forests in southwestern Cambodia: I. Plant biomass". In: *Nature and Life in Southeast Asia* 6, pp. 1–51. URL: http://www.publikationen.ub.uni-frankfurt.de/frontdoor/index/index/docId/27798 (visited on 2017-07-21).

Huxley, J. S. and Teissier, G. (1936). "Terminology of relative growth". In: *Nature* 137, pp. 780–781. DOI: 10.1038/137780b0.

IPCC (2000). *Land use, land-use change and forestry. A special report of the Intergovernmental Panel on Climate Change*. Cambridge University Press. ISBN: 978-0-521-80495-0.

Jarque, C. M. and Bera, A. K. (1980). "Efficient tests for normality, homoscedasticity and serial independence of regression residuals". In: *Economics Letters* 6 (3), pp. 255–259. DOI: 10.1016/0165-1765(80)90024-5.

Jasechko, S., Sharp, Z. D., Gibson, J. J., Yi, Y., and Fawcett, P. J. (2013). "Terrestrial water fluxes dominated by transpiration". In: *Nature* 496 (7445), pp. 347–350. DOI: 10.1038/nature11983.

Jelalian, A. V. (1992). *Laser radar systems*. Artech House. ISBN: 978-0-89006-554-9.

Jupp, D. L. B., Culvenor, D. S., Lovell, J. L., Newnham, G. J., Strahler, A. H., et al. (2009). "Estimating forest LAI profiles and structural parameters using a ground-based laser called Echidna". In: *Tree Physiology* 29 (2), pp. 171–181. DOI: 10.1093/treephys/tpn022.

Jupp, D. L. B. and Lovell, J. L. (2007). *Airborne and ground-based LiDAR systems for forest measurement: background and principles*. CSIRO Marine and Atmospheric Research. ISBN: 978-1-921232-68-8.

Kaasalainen, S., Jaakkola, A., Kaasalainen, M., Krooks, A., and Kukko, A. (2011). "Analysis of incidence angle and distance effects on terrestrial laser scanner intensity: search for correction methods". In: *Remote Sensing* 3 (10), pp. 2207–2221. DOI: 10.3390/rs3102207.

Kang, Z., Li, J., Zhang, L., Zhao, Q., and Zlatanova, S. (2009). "Automatic registration of terrestrial laser scanning point clouds using panoramic reflectance images". In: *Sensors* 9 (4), pp. 2621–2646. DOI: 10.3390/s90402621.

Ketterings, Q. M., Coe, R., Noordwijk, M. van, Ambagau', Y., and Palm, C. A. (2001). "Reducing uncertainty in the use of allometric biomass equations for predicting above-ground tree biomass in mixed secondary forests". In: *Forest Ecology and Management* 146 (1), pp. 199–209. DOI: 10.1016/S0378-1127(00)00460-6.

Kittredge, J. (1944). "Estimation of the amount of foliage of trees and stands". In: *Journal of Forestry* 42 (12), pp. 905–912. URL: http://ingentaconnect.com/contentone/saf/jof/1944/00000042/00000012/art00013 (visited on 2017-07-21).

Knapic, S., Louzada, J. L., Leal, S., and Pereira, H. (2008). "Within-tree and between-tree variation of wood density components in cork oak trees in two sites in Portugal". In: *Forestry* 81 (4), pp. 465–473. DOI: 10.1093/forestry/cpn012.

Koenker, R. (2005). *Quantile regression*. Cambridge Univeristy Press. ISBN: 978-0-521-84573-1.

Koenker, R. and Park, B. J. (1996). "An interior point algorithm for nonlinear quantile regression". In: *Journal of Econometrics* 71 (1), pp. 265–283. DOI: 10.1016/0304-4076(96)84507-6.

Kozlowski, J. and Konarzewski, M. (2004). "Is West, Brown and Enquist's model of allometric scaling mathematically correct and biologically relevant?" In: *Functional Ecology* 18 (2), pp. 283–289. DOI: 10.1111/j.0269-8463.2004.00830.x.

Landau, L. D. and Lifshitz, E. M. (1987). *Fluid Mechanics. Course of Theoretical Physics, Volume 6*. Pergamon Press. ISBN: 978-0-08-033933-7.

Landi, G. (2002). "Properties of the centre of gravity as an algorithm for position measurements". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 485 (3), pp. 698–719. DOI: 10.1016/S0168-9002(01)02071-X.

Lebedys, A. and Li, Y. (2014). *Contribution of the forestry sector to national economies, 1990-2011*. Food and Agriculture Organization of the United Nations. URL: http://www.fao.org/publications/card/en/c/0c077344-03b5-4d0b-9953-4606c27e1884 (visited on 2017-07-21).

Lee, S. S. and Yahya, N. S. (1999). "Fungi associated with heart rot of Acacia Mangium trees in Peninsular Malaysia and East Kalimantan". In: *Journal of Tropical Forest Science* 11 (1), pp. 240–254. URL: https://jstor.org/stable/43582526 (visited on 2017-07-21).

Lefsky, M. A., Harding, D. J., Keller, M., Cohen, W. B., Carabajal, C. C., et al. (2005). "Estimates of forest canopy height and aboveground biomass using ICESat". In: *Geophysical Research Letters* 32 (22). DOI: 10.1029/2005GL023971.

Lewis, P. (1999). "Three-dimensional plant modelling for remote sensing simulation studies using the Botanical Plant Modelling System". In: *Agronomie* 19 (34), pp. 185–210. DOI: 10.1051/agro:19990302.

Lewis, S. L., Sonké, B., Sunderland, T., Begne, S. K., Lopez-Gonzalez, G., et al. (2013). "Above-ground biomass and structure of 260 African tropical forests". In: *Philosophical Transactions of the Royal Society of London B: Biological Sciences* 368 (1625). DOI: 10.1098/rstb.2012.0295.

Li, Z., Jupp, D. L. B., Strahler, A. H., Schaaf, C. B., Howe, G., et al. (2016). "Radiometric calibration of a dual-wavelength, full-waveform terrestrial LiDAR". In: *Sensors* 16 (3). DOI: 10.3390/s16030313.

Lim, K., Treitz, P., Wulder, M., St-Onge, B., and Flood, M. (2003). "LiDAR remote sensing of forest structure". In: *Progress in Physical Geography* 27 (1), pp. 88–106. DOI: 10.1191/0309133303pp360ra.

Lovell, H. and MacKenzie, D. (2016). "Allometric equations and timber markets: an important forerunner of REDD+?" In: *The politics of carbon markets*. Routledge. ISBN: 978-1-138-20515-4.

Lovell, J. L., Jupp, D. L. B., Culvenor, D. S., and Coops, N. C. (2003). "Using airborne and ground-based ranging LiDAR to measure canopy structure in Australian forests". In: *Canadian Journal of Remote Sensing* 29 (5), pp. 607–622. DOI: 10.5589/m03-026.

Lukács, G., Martin, R., and Marshall, D. (1998). "Faithful least-squares fitting of spheres, cylinders, cones and tori for reliable segmentation". In: *Computer Vision — ECCV'98, V, Freiburg, Germany, June 2–6*. DOI: 10.1007/BFb0055697.

Maas, H.-G., Bienert, A., Scheller, S., and Keane, E. (2008). "Automatic forest inventory parameter determination from terrestrial laser scanner data". In: *International Journal of Remote Sensing* 29 (5), pp. 1579–1593. DOI: 10.1080/01431160701736406.

Malhi, Y., Phillips, O. L., Lloyd, J., Baker, T., Wright, J., et al. (2002). "An international network to monitor the structure, composition and dynamics of Amazonian forests (RAINFOR)". In: *Journal of Vegetation Science* 13 (3), pp. 439–450. DOI: 10.1111/j.1654-1103.2002.tb02068.x.

Marquardt, D. W. (1963). "An algorithm for least-Squares estimation of nonlinear parameters". In: *Journal of the Society for Industrial and Applied Mathematics* 11 (2), pp. 431–441. DOI: 10.1137/0111030.

Marthews, T. R., Riutta, T., Oliveras Menor, I., Urrutia, R., Moore, S., et al. (2014). *Measuring tropical forest carbon allocation and cycling: a RAINFOR-GEM field manual for intensive census plots (v3.0)*. Global Ecosystems Monitoring network. URL: http://gem.tropicalforests.ox.ac.uk (visited on 2017-07-21).

Mascaro, J., Litton, C. M., Hughes, R. F., Uowolo, A., and Schnitzer, S. A. (2011). "Minimizing bias in biomass allometry: model selection and log-transformation of

data". In: *Biotropica* 43 (6), pp. 649–653. DOI: 10.1111/j.1744-7429.2011.00798.x.

Mitchard, E. T. A., Feldpausch, T. R., Brienen, R. J. W., Lopez-Gonzalez, G., Monteagudo, A., et al. (2014). "Markedly divergent estimates of Amazon forest carbon density from ground plots and satellites". In: *Global Ecology and Biogeography* 23 (8), pp. 935–946. DOI: 10.1111/geb.12168.

Nambiar, K. R. (2004). *Lasers: principles, types and applications*. New Age International. ISBN: 978-81-224-1492-9.

Newnham, G. J., Armston, J. D., Calders, K., Disney, M. I., Lovell, J. L., et al. (2015). "Terrestrial laser scanning for plot-scale forest measurement". In: *Current Forestry Reports* 1 (4), pp. 239–251. DOI: 10.1007/s40725-015-0025-5.

Neyman, J. and Scott, E. L. (1960). "Correction for bias introduced by a transformation of variables". In: *The Annals of Mathematical Statistics* 31 (3), pp. 643–655. DOI: 10.1214/aoms/1177705791.

Nock, C. A., Geihofer, D., Grabner, M., Baker, P. J., Bunyavejchewin, S., et al. (2009). "Wood density and its radial variation in six canopy tree species differing in shade-tolerance in western Thailand". In: *Annals of Botany* 104 (2), pp. 297–306. DOI: 10.1093/aob/mcp118.

Packard, G. C., Birchard, G. F., and Boardman, T. J. (2011). "Fitting statistical models in bivariate allometry". In: *Biological Reviews* 86 (3), pp. 549–563. DOI: 10.1111/j.1469-185X.2010.00160.x.

Paine, C. E. T., Stahl, C., Courtois, E. A., Patiño, S., Sarmiento, C., et al. (2010). "Functional explanations for variation in bark thickness in tropical rain forest trees". In: *Functional Ecology* 24 (6), pp. 1202–1210. DOI: 10.1111/j.1365-2435.2010.01736.x.

Pan, Y., Birdsey, R. A., Fang, J., Houghton, R., Kauppi, P. E., et al. (2011). "A large and persistent carbon sink in the world's forests". In: *Science* 333 (6405), pp. 988–993. DOI: 10.1126/science.1201609.

Pasquis, R. and Bouamrane, M. (2004). "Deforestation and its impact on biological diversity: a question of scale". In: *Beyond tropical deforestation: from tropical deforestation to forest cover dynamics and forest development*. Centre de Coopération

Internationale en Recherche Agronomique pour le Développement, United Nations Educational, Scientific, and Cultural Organization. ISBN: 978-92-3-103941-6.

Pauly, M., Keiser, R., and Gross, M. (2003). "Multi-scale feature extraction on point-sampled surfaces". In: *Computer Graphics Forum* 22 (3), pp. 281–289. DOI: 10.1111/1467-8659.00675.

Peres, C. A., Barlow, J., and Laurance, W. F. (2006). "Detecting anthropogenic disturbance in tropical forests". In: *Trends in Ecology and Evolution* 21 (5), pp. 227–229. DOI: 10.1016/j.tree.2006.03.007.

Pfeifer, N., Hofle, B., Briese, C., and Haring, A. (2008). "Analysis of the backscattered energy in terrestrial laser scanning data". In: *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XXXVII, Beijing, China, July 7–11*. URL: http://www.isprs.org/proceedings/XXXVII/congress/5_pdf/180.pdf (visited on 2017-07-21).

Pfennigbauer, M. and Ullrich, A. (2010). "Improving quality of laser scanning data acquisition through calibrated amplitude and pulse deviation measurement". In: *Proceedings of SPIE 7684, Laser Radar Technology and Applications, XV, Florida, United States, April 05*. DOI: 10.1117/12.849641.

Picard, N., Saint-André, L., and Henry, M. (2012). *Manual for building tree volume and biomass allometric equations: from field measurement to prediction*. Centre de Coopération Internationale en Recherche Agronomique pour le Développement, Food, and Agriculture Organization of the United Nations. ISBN: 978-92-5-107347-6.

Pypker, T. G., Bond, B. J., Link, T. E., Marks, D., and Unsworth, M. H. (2005). "The importance of canopy structure in controlling the interception loss of rainfall: examples from a young and an old-growth Douglas-fir forest". In: *Agricultural and Forest Meteorology* 130 (1–2), pp. 113–129. DOI: 10.1016/j.agrformet.2005.03.003.

Rabbani, T., Heuvel, F. A. van den, and Vosselmann, G. (2006). "Segmentation of point clouds using smoothness constraint". In: *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, XXXVI, Dresden, Germany, September 25–27*. URL: http://www.isprs.org/proceedings/XXXVI/part5/paper/RABB_639.pdf (visited on 2017-07-21).

Raumonen, P., Casella, E., Calders, K., Murphy, S., Åkerblom, M., et al. (2015). "Massive-scale tree modelling from TLS data". In: *International Archives of Photogramme-*

try, Remote Sensing and Spatial Information Sciences, II-3/W4, Munich, Germany, March 25–27. DOI: 10.5194/isprsannals-II-3-W4-189-2015.

Raumonen, P., Kaasalainen, M., Åkerblom, M., Kaasalainen, S., Kaartinen, H., et al. (2013). "Fast automatic precision tree models from terrestrial laser scanner data". In: *Remote Sensing* 5 (2), pp. 491–520. DOI: 10.3390/rs5020491.

Richards, P. W. (1952). *The tropical rain forest: an ecological study*. Cambridge University Press. ISBN: 978-0-521-29658-8.

RIEGL Laser Measurement Systems GmbH (2016). *RiSCAN Pro*. Version 2.1. URL: http://www.riegl.com/products/software-packages/riscan-pro (visited on 2017-07-21).

Rusu, R. B. and Cousins, S. (2011). "3D is here: Point Cloud Library (PCL)". In: *IEEE International Conference on Robotics and Automation, Shanghai, China, May 9–13*. DOI: 10.1109/ICRA.2011.5980567.

Saatchi, S. S., Harris, N. L., Brown, S., Lefsky, M., Mitchard, E. T. A., et al. (2011). "Benchmark map of forest carbon stocks in tropical regions across three continents". In: *Proceedings of the National Academy of Sciences* 108 (24), p. 9899. DOI: 10.1073/pnas.1019576108.

Schaepman-Strub, G., Schaepman, M. E., Painter, T. H., Dangel, S., and Martonchik, J. V. (2006). "Reflectance quantities in optical remote sensing - definitions and case studies". In: *Remote Sensing of Environment* 103 (1), pp. 27–42. DOI: 10.1016/j.rse.2006.03.002.

Schnabel, R., Wahl, R., and Klein, R. (2007). "Efficient RANSAC for point-cloud shape detection". In: *Computer Graphics Forum* 26 (2), pp. 214–226. DOI: 10.1111/j.1467-8659.2007.01016.x.

Schofield, L. A., Danson, F. M., Entwistle, N. S., Gaulton, R., and Hancock, S. (2016). "Radiometric calibration of a dual-wavelength terrestrial laser scanner using neural networks". In: *Remote Sensing Letters* 7 (4), pp. 299–308. DOI: 10.1080/2150704X.2015.1134843.

Shan, J. and Toth, C. K. (2008). *Topographic laser ranging and scanning: principles and processing*. CRC Press. ISBN: 978-1-4200-5142-1.

Slik, J. W. F., Arroyo-Rodriguez, V., Aiba, S.-I., Alvarez-Loayza, P., Alves, L. F., et al. (2015). "An estimate of the number of tropical tree species". In: *Proceedings of the National Academy of Sciences* 112 (24). DOI: 10.1073/pnas.1423147112.

Slik, J. W. F., Paoli, G., McGuire, K., Amaral, I., Barroso, J., et al. (2013). "Large trees drive forest aboveground biomass variation in moist lowland forests across the tropics". In: *Global Ecology and Biogeography* 22 (12), pp. 1261–1271. DOI: 10.1111/geb.12092.

Smith, A., Astrup, R., Raumonen, P., Liski, J., Krooks, A., et al. (2014). "Tree root system characterization and volume estimation by terrestrial laser scanning and quantitative structure modeling". In: *Forests* 5 (12), pp. 3274–3294. DOI: 10.3390/f5123274.

Sneed, E. D. and Folk, R. L. (1958). "Pebbles in the Lower Colorado River, Texas a Study in Particle Morphogenesis". In: *The Journal of Geology* 66 (2), pp. 114–150. URL: http://www.jstor.org/stable/30058239 (visited on 2017-07-21).

Tansey, K., Selmes, N., A., A., Tate, N. J., and Denniss, A. (2009). "Estimating tree and stand variables in a Corsican Pine woodland from terrestrial laser scanner data". In: *International Journal of Remote Sensing* 30 (19), pp. 5195–5209. DOI: 10.1080/01431160902882587.

The World Bank (2004). *Sustaining forests: a development strategy.* The International Bank for Reconstruction and Development / The World Bank. ISBN: 978-0-8213-5755-2.

Thies, M., Pfeifer, N., Winterhalder, D., and Gorte, B. G. H. (2004). "Three-dimensional reconstruction of stems for assessment of taper, sweep and lean based on laser scanning of standing trees". In: *Scandinavian Journal of Forest Research* 19 (6), pp. 571–581. DOI: 10.1080/02827580410019562.

Utts, J. M. (1982). "The rainbow test for lack of fit in regression". In: *Communications in Statistics - Theory and Methods* 11 (24), pp. 2801–2815. DOI: 10.1080/03610928208828423.

Vastaranta, M., Melkas, T., Holopainen, M., Kaartinen, H., Hyyppä, J., et al. (2009). "Laser-based field measurements in tree-level forest data acquisition". In: *The Photogrammetric Journal of Finland* 21 (2), pp. 51–61. URL: https://foto.aalto.fi/

seura/julkaisut/pjf/pjf_e/2009/Vastaranta_et_al_2009_PJF.pdf (visited on 2017-07-21).

Verroust, A. and Lazarus, F. (2000). "Extracting skeletal curves from 3D scattered data". In: *The Visual Computer* 16 (1), pp. 15–25. DOI: 10.1007/PL00007210.

Wagner, W., Ullrich, A., Ducic, V., Melzer, T., and Studnicka, N. (2006). "Gaussian decomposition and calibration of a novel small-footprint full-waveform digitising airborne laser scanner". In: *ISPRS Journal of Photogrammetry and Remote Sensing* 60 (2), pp. 100–112. DOI: 10.1016/j.isprsjprs.2005.12.001.

Watt, P. J. and M., D. D. N. (2005). "Measuring forest structure with terrestrial laser scanning". In: *International Journal of Remote Sensing* 26 (7), pp. 1437–1446. DOI: 10.1080/01431160512331337961.

Weinmann, M., Weinmann, M., Hinz, S., and Jutzi, B. (2011). "Fast and automatic image-based registration of TLS data". In: *ISPRS Journal of Photogrammetry and Remote Sensing* 66 (6), pp. 62–70. DOI: 10.1016/j.isprsjprs.2011.09.010.

Welles, J. M. and Cohen, S. (1996). "Canopy structure measurement by gap fraction analysis using commercial instrumentation". In: *Journal of Experimental Botany* 47 (9), pp. 1335–1342. DOI: 10.1093/jxb/47.9.1335.

West, G. B., Brown, J. H., and Enquist, B. J. (1997). "A general model for the origin of allometric scaling laws in biology". In: *Science* 276 (5309), pp. 122–126. DOI: 10.1126/science.276.5309.122.

White, H. (1980). "A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity". In: *Econometrica* 48 (4), pp. 817–838. DOI: 10.2307/1912934.

Widlowski, J.-L., Pinty, B., Lavergne, T., Verstraete, M. M., and Gobron, N. (2005). "Using 1-D models to interpret the reflectance anisotropy of 3-D canopy targets: issues and caveats". In: *IEEE Transactions on Geoscience and Remote Sensing* 43 (9), pp. 2008–2017. DOI: 10.1109/TGRS.2005.853718.

Williams, M. (2003). *Deforesting the earth: from prehistory to global crisis*. University of Chicago Press. ISBN: 978-0-226-89926-8.

Woodhouse, I. H., Mitchard, E. T. A., Brolly, M., Maniatis, D., and Ryan, C. M. (2012). "Radar backscatter is not a 'direct measure' of forest biomass". In: *Nature Climate Change* 2 (8), pp. 556–557. DOI: 10.1038/nclimate1601.

Wu, C. F. J. (1986). "Jackknife, bootstrap and other resampling methods in regression analysis". In: *The Annals of Statistics* 14 (4), pp. 1261–1295. DOI: 10.1214/aos/1176350142.

Xiao, X., White, E. P., Hooten, M. B., and Durham, S. L. (2011). "On the use of log-transformation vs. nonlinear regression for analyzing biological power laws". In: *Ecology* 92 (10), pp. 1887–1894. DOI: 10.1890/11-0538.1.

Zianis, D. and Radoglou, K. (2006). "Comparison between empirical and theoretical biomass allometric models and statistical implications for stem volume predictions". In: *Forestry* 79 (4), pp. 477–487. DOI: 10.1093/forestry/cpl028.